DOCUMENT RESUME

ED 427 649 IR 019 168

AUTHOR	Maurer, Hermann, Ed.
TITLE	WebNet 96 Conference Proceedings (San Francisco, California, October 15-19, 1996).
INSTITUTION	Association for the Advancement of Computing in Education, Charlottesville, VA.
PUB DATE	1996-10-00
NOTE	930p.; For selected individual papers, see IR 019 169-198. Many figures and tables are illegible.
AVAILABLE FROM	Web site: http://aace.virginia.edu/aace/conf/webnet/proc96.html; also archived on WebNet 98 CD-ROM (includes 1996, 1997, 1998) AACE Membership/CD orders, P.O. Box 2966, Charlottesville, VA 22902; Fax: 804-978-7449 ($35, AACE members, $40, nonmembers).
PUB TYPE	Collected Works - Proceedings (021)
EDRS PRICE	MF06/PC38 Plus Postage.
DESCRIPTORS	Access to Information; Authoring Aids (Programming); Computer Science; Computer Software; Courseware; Databases; Distance Education; Educational Media; Educational Strategies; *Educational Technology; Electronic Libraries; Elementary Secondary Education; *Hypermedia; Information Technology; Instructional Design; Multimedia Materials; Postsecondary Education; *World Wide Web
IDENTIFIERS	Electronic Commerce; Software Tools; Virtual Classrooms; *Web Sites

ABSTRACT

This proceedings contains 80 full papers, 12 posters/demonstrations, 108 short papers, one panel, and one tutorial, all focusing on World Wide Web applications. Topics include: designing hypertext navigation tools; Web site design; distance education via the Web; instructional design; the world-wide market and censorship on the Web; customer support via the Web; VRML; tracking Web users; developing adaptive electronic textbooks on the Web; developing secure Web accessible databases; automating hypermedia course creation and maintenance; innovative resources for educational and public information; interactive justice trough the Web; 3D visualization of Web semantic content for browsing and query formation; Web authoring tools; the impact of the Web on education; multilingual font service; networked learning environments; Web conceptual space; the impact of Web publication on a child writer's development; concurrent manufacturing on the Web; browsing; virtual office hours; designing digital libraries for post-literate patrons; electronic commerce; a distance education how-to manual; media, multimedia, and language learning; managing the virtual classroom; a course in multimedia technology for non-computer science majors; full-text retrieval; collaborative learning/teaching on the Web: the web as a teaching tool; computer mediated communication; active messaging in Java; an electronic shopping agent; helping teachers use anti-virus Web sites; teaching personally sensitive material over the Internet; global distance learning and the transformation of knowledge construction; gender and the human computer user interface; school participation in the development of commercial, Web-based curriculum resources; a virtual campus on the Web; authentic activities and the Internet; a tool for creating and browsing in

+++++ ED427649 Has Multi-page SFR---Level=1 +++++

case libraries; providing real-time instruction on the Web; discursive networking; Web language teaching; and how to set up and maintain an image database on the Web. (DLS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 427 649

WebNet 96 Conference Proceedings
San Francisco, CA
October 15-19, 1996

Edited by Hermann Maurer

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

Published by Association for the Advancement of Computing in Education

IR019168

WEBNET 96

Table of Contents

FULL PAPERS

BRIAN BEVIRT
Designing hypertext navigation tools

MARTIN BICHLER & STEFAN NUSSER
Developing structured WWW-sites with W3DT

JOHN BIGELOW
Developing an Internet section of a management course: Transporting learning Premises Across Media

MICHAEL BJORN & YIN YUE CHEN
The world-wide market: Living with the realities of censorship on the Internet

MANFRED BOGEN, MICHAEL LENZ & SUSANNE ZIER
A Broadcasting Company goes Internet

WILLIAM A. BOGLEY, JON DORBOLO, ROBBERT O.ROBSON & JOHN A. SECHREST
New pedagogies and tools for Web-based calculus

ELLEN BORKOWSKI, DAVID HENRY, LIDA LARSEN & DEBORAH MATEIK
Supporting teaching and learning via the Web: Transforming hardcopy linear mindsets into Web flexible creative thinking

ALEXANDRA BOST
The WWW as a primary source of customer support

WOLFGANG BROLL
VRML and the Web: A fundament for multi-user virtual environments on the Internet

CHRIS BROWN & STEVE BENFORD
Tracking WWW Users: Experience from the Design of HyperVis

PETER BRUSILOVSKY, ELMAR SCHWARZ & GERHARD WEBER
A tool for developing adaptive electronic textbooks on WWW

ROBERT BUCCIGROSSI, ALBERT CROWLEY & DANIEL TURNER
A comprehensive system to develop secure Web accessible databases

ANTONIO CAPANI & GABRIEL DE DOMINICIS
Web Algebra

SCOTT CAPDEVIELLE (BUSINESS/CORPORATE)
Capturing the state of the web

CURTIS A. CARVER & CLARK RAY
Automating hypermedia course creation and maintenance

LEE LI-JEN CHEN & BRIAN R. GAINES
Methodological issues in studying and supporting awareness on the World Wide Web

CAROL A. CHRISTIAN
Innovative resources for educational and public information: Electronic services, data and information from NASA's Hubble Space Telescope and other NASA missions

BETTY COLLIS, TOINE ANDERNACH & NICO VAN DIEPEN
The Web as process tool and product environment for group-based project work in higher education

MARGARET CORBIT & MICHAEL HERZOG
Explorations: The Cornell Theory Center's online science book

CHANTAL D'HALLUIN, STEPHANE RETHORE, BRUNO VANHILLE & CLAUDE VIEVILLE
Designing a course on the Web: The point of view of a training institute

P.M.D. DE BRA
Teaching hypertext and hypermedia through the Web

DIETER W. FELLNER & OLIVER JUCKNATH
MRTspace: Multi-user 3D environments using VRML

BARRY FENN & JENNY SHEARER
Delivering the daily Us

RICHARD H. FOWLER, WENDY A.L. FOWLER & JORGE L. WILLIAMS
3D visualization of WWW semantic content for browsing and query formulation

G. FULANTELLI & M. ALLEGRA
Interactive justice through the Web

BRIAN R. GAINES (INVITED SPEAKER)
Convergence to the Information Superhighway

BRIAN R. GAINES, MILDRED L.G. SHAW & LEE LI-JEN CHEN
Utility, usability and likeability: Dimensions of the Net and Web

SUSAN GAUCH & GUIJUN WANG
Information fusion with ProFusion

ROBIN E. GOODMAN & PAUL S. CHINOWSKY
The Web in interdisciplinary design education

CHETAN GOPAL & CESAR A.C. TEIXEIRA
MHEG-5 delivery in the DAVIC environment

A. HASHMI, B. BARRY, R. LUNDQUIST & M. KRISHNAMOORTHY
Aster servers: How to make Aster more widely available

ERIK P. HAUGSJAA
Collaborative learning through a knowledge-based WWW authoring tool

CAROLINE HECHT & KATHY BARBIERI
Math and Science Gateway - A successful model for high school educators and students

DANIEL J. HELM, RAYMOND J. D'AMORE & PUCK-FAI YAN
MITRE information discovery system

D.J. HOBBS & R.J. TAYLOR
The impact on education of the World Wide Web

JENS F. JENSEN
Mapping the Web: A Media Typology for Information Traffic Patterns on the Internet Highway

W. LEWIS JOHNSON, TYLER BLAKE & ERIN SHAW
Automated management and delivery of distance courseware

R. JOHNSON, J. DURRANCE, J. KOZYRA, C. ALEXANDER, T. CLARKE, J. LINKER, D. MASTIE, J. GREEN, P. ORSELLI & C. RASMUSSEN
Windows to the universe

JIM JUNGBAUER, PATRICIA BAGGETT & ANDRZEJ EHRENFEUCHT
A browsing system based on multimedia cohesion

WING KAY KAN, STEPHEN S.M. WONG, VINCENT S.Y. YIU, GILBERT H.F. YOUNG, HAK WAI CHAN, HENRY W.H. CHEUNG
On-demand multi-lingual font service on heterogeneous computer platforms

JUDY KEINER
Real audiences-worldwide: A case study of the impact of WWW publication on a child writer's development

ROBERT E. KENT & CHRISTIAN NEUSS
Web conceptual space

ARKADI KOSMYNIN
An information broker for adaptive distributed resource discovery service

ANASTASIOS KOUTOUMANOS, NIKOLAOS PAPASPYROU, SIMEON RETALIS & EMMANOUEL SKORDALAKIS
Towards a novel networked learning environment

ROB KREMER & BRIAN R. GAINES
Embedded interactive concept maps in Web documents

JAMES M. LAFFEY & DALE MUSSER
Building Internet-based electronic performance support for teaching and learning

ANDREW Z. LAPSLEY, BRIAN R. GAINES & DOUGLAS H. NORRIE
Concurrent manufacturing on the Web

GERRY LEBING, MAURA HOGAN & WESLEY GURNEY (BUSINESS/CORPORATE)
On-the-Fly Web Pages Minus the HTML Files

S. LENMAN, HENRY SEE, MICHAEL CENTURY, & BRUCE PENNYCOOK
Merz: Creating personal and shared spaces on the World Wide Web

J. LENNON & H. MAURER
Aspects of large World Wide Web systems

THIERRY LEVOIR & REMI BASTIDE
SADS: Generic Web access to space data

PAUL LOSLEBEN, ALICE S. TULL, & ERMING ZHOU
Help Exchange: An arbitrated system for a help network

YANNICK MARCHAND, JEAN-LUC GUERIN & JEAN-PAUL BARTHES
From a set of technical documents to a hypertext system on the Web

J.M. MARTINEZ & F. MORAN
Catalog: A WWW gateway for DBMSs

HERMANN MAURER, KEITH ANDREWS & NICK SCHERBAKOV
Browsing hypermedia composites: An algebraic approach

SUZANNE MCMAHON
Internet India

CRAIG A. MERLIC & MATTHEW J. WALKER
Virtual office hours: A communication tool for students and teachers

GREG MILLBANK
Using Java-Based Controlled Process Simulation for Training via Intranet or Internet

KIYOSHI NAKABAYASHI, MINA MARUYAMA, YOSHIMASA KOIKE, YOSHIMI FUKUHARA & YUKIHIRO NAKAMURA
An intelligent tutoring system on the WWW supporting interactive simulation environment with a multimedia viewer control mechanism

PETER J. NURNBERG, ERICH R. SCHNEIDER & JOHN J. LEGGETT
Designing digital libraries for post-literate patrons

ANDREW ODLYZKO (INVITED SPEAKERS)
The Bumpy Road of Electronic Commerce

NIKOLAOS PAPASPYROU, ANASTASIOS KOUTOUMANOS, HERMANN MAURER & EMMANOUEL SKORDALAKIS
An Ongoing experiment in ODL using new technologies

ANGIE PARKER
A distance education how-to manual: Recommendations from the field

RICHARD PONIGHAUS & JOHANN MITLOHNER

Relational Databases and the World Wide Web: Automatic Generation of Hypertext based on Reverse-engineered meta information

SARAH PRICE, PATRICK MACANDREW, ROGER RIST, TERRY MAYES, ERIC BONHARME, RAY LAND, MARY CUTTLE, JEFF HAYWOOD & HAMISH MACLEOD
The MARBLE Project: A collaborative framework for educational courseware design

CLARK N. QUINN (BUSINESS/CORPORATE)
Open Net and Australian On-line Education Market

JUDY COSSEL RICE, PAUL F. MERRILL & CAROL LEE HAWKINS
Procedures for Creating Useful Web Sites

KLAUS SCHMARANZ
Professional electronic publishing in Hyper-G: The next generation publishing solution on the Web

MILDRED L.G. SHAW & BRIAN R. GAINES
WebGrid: Knowledge elicitation and modeling on the Web

PETER SHEA
Media multimedia and language learning: A review of the literature

JENNY SHEARER
Mediation in the Internet

JUDI R.THOMSON, JOHN E. COOKE & JIM E. GREER
The MicroWeb toolkit: Bringing the WWW to the classroom

MATSUMOTO TSUTOMU, TOMONORI WASAKI, MITOMO OKAYAMA, KATSUNORI HORIKAWA, TOSHIRIRO HAYASHI, & MIKE NGA
A Web site for counseling on students' future courses

MARK VAN GORP & PETER BOYSEN
ClassNet: Managing the virtual classroom

R.P. VAN DE RIET & J.F.M. BURG
Modelling alter egos in cyberspace: Who is responsible?

COGNITION AND TECHNOLOGY GROUP AT VANDERBILT (CTGV)
A SMART curriculum for middle-school science instruction: A Web-based curriculum integrating assessment & instruction

JERRY WAXMAN
A course in Multimedia technology for non-computer science majors

MARTIN D. WESTHEAD
EPIC: Building a structured learning environment

IAN H. WITTEN, CRAIG G. NEVILL-MANNING AND SALLY JO CUNNINGHAM (INVITED SPEAKER)
Digital Libraries Based on Full-Text Retrieval

ZHENGMAI ZHAO & JIAN HU
Supporting provision and access of educational visual resources on the web

POSTERS/DEMONSTRATIONS

JORN BOLLMEYER
psrohg-A tool for integrating presentations into hypermedia systems

ISABEL BORRAS
Development and evaluation of World Wide Web courseware for teacher education

JOSEPH BOWMAN, JR., ALIYA HOLMES, DONOVAN SMITH, MARC CALLENDAR, ETWIN MAPP-BOWMAN
Science Technology entry program: Integrating technology into a cross curricula design

JOHN DIETZ
The Instructional Designer community Center and Library

DAVIS KUNTZ & NED WALTHALL
RIDE: A tool for developing web-based instruction and assessment

DEBORAH MCALLISTER & WENDY K. BARTNIK
Preservice teacher education and the Web: Shifting paradigms, new partnerships

KIYOTERU MOTODA, WAKAKO TOKUMARU , HIDEO KIYOHARA & MATSUMOTO TSUTOMU
Evaluation of multimedia authoring tool TSUMIKI through developing AIDS courseware

TERRI J. NELSON & WALTER CARL OLIVER
Un Meurtre a Cinet (Un homicidio en Toluca): An e-mail whodunnit to develop writing competence in intermediate level language classes offered at a distance

JAMES PARADISO
A model-in-progress for designing web-based assignments in business management education

JOAN ROBERTSON, JOHN SUNDSEN, & JOHN BOLLES
The Human Brain: A frames based image oriented quiz

HARDY SCHULZE, S. BORNSEIN & THOMAS BAEHRING
The World Wide Web: A new approach in Public Health Education and risk awareness of diabetes at

NANCY WYATT, SUSAN WARE & MATTHEW BODEK
Collaborative learning/teaching on the Web: Lessons

SHORT PAPERS

MARCO AIELLO & MARTIN HAEBERLI
Integrating servers for communication over the Internet

GREGORY ALOIA, RODNEY RIEGLE, & PERRY SCHOON
The web as a teaching tool

SHANNON ALPERT
Bringing instructional design principles on-line: An instructional technologists

approach to Web development

CHRIS ANDERSON, CHRIS BURNS & MICHAEL KLEMME
Watermarking for the HyperWave Hypermedia System

PAMELA ANDREW AND GRAEME HART
Impact of the Introduction of Web-Based Teaching on Academic Staff

MASAYASU AOTANI
Distance education and the Web in engineering and mathematics: Examples and projects at Stanford and UC Berkely

MASAYASU AOTANI
Japanese instruction over the Web: Examples at MIT and UC Berkeley

BASSAM AOUN
Robots And The Web: An Electronic Shopping Agent

ALIREZA BAHREMAN & NICK ZHANG
Active Messaging in Java

BRENDA BANNAN & WILLIAM D. MILHEIM
Design, Development and Delivery of Instructional Materials over the Internet

LAURA BARTOLO
ALCOM/NIST Phase Separation Project

ISABEL BORRAS
Internet for teachers: Case study of a graduate course

JOSEPH BOWMAN, JR.
Creating Educational Webmasters: Cyberhood Cooperative Learning Project

ROBERT G. BRANNAN & MARY MCPHAIL
Creating Web-pages in education on a limited budget: Strategies for success in a cooperative extension

MARC H. BROWN & WILLIAM E. WEIHL
Zippers: A Focus+Context Display of Web Pages

AGATA BRUSEGAN, ALVISE DE MICHELIS, LUIGI GIACULLI & LUISELLA ROMEO
The VENetIan Virtual Archive Project and a new approach to cultural heritage

ERIC CHEUNG
Cyberschool- Singapore Polytechnic Model

H. MICHAEL CHUNG
Information Superhighway: Architectures and applications

ALEX CUTHBERT
Collaborative design: Cognitive approach to information resources

ALANIS DAVID & RAMSES HERRERA
A Web based decision support system

HOLLY DEVAUL & NANCY BUTLER SONGER

Kids as global scientists: Promoting a web of collaboration through Internet-based curricula

SJOERD DE VRIES & HENK DONKER
HERODOTUS: An Educational Site as an Integral Part of a Study Book

GREG DIXON
Knowledge Explorer Centre: The Web as multimedia publishing medium

PHILIP DUCHASTEL & SUE SPAHN
Design for Web-based learning

ERIK DUVAL &HENK OLIVIE
The Web is not a distributed multimedia system (yet?)

ED DZIERZAK, KENT HAYES & WILLIAM LUCAS
WEB-EDD: Electronic Document delivery via World Wide Web

JOHN EKLUND & JAMES SAWERS
Customising Web-based course delivery in WEST with adaptive navigation support

ERNEST B. EZELL, JR. & DIANE K. FREY
An innovative teaching technique using the Web to deliver case study instruction

KEVIN C. FACEMYER, GARY BROWN & NILS PETERSON
Developments in Internet-based community building

JEFFREY FERGUSON
Introduction and description of the Interactive Image Format (.iif) for the World Wide Web

RICHARD FERGUSON
The WWW in distance education: A new technology for some old problems

TIMOTHY FOSTER & MARILYN FRITZ
AHP: The Academic Home Page

LYNN FOX
Helping teachers use anti-virus sites on the World Wide Web

JOAN FREEDMAN & HAROLD LEHMANN
Providing medical education resources over the World Wide Web

MOLLY FREEMAN
Global distance learning and the transformation of knowledge construction

NICOLA FREGA, ANTONIO VOLPENTESTA, TOMMASO FASANO & ANTONIO GRECO
Homeland: An Integrated Set of Tools to Support Regional telecommunications

JAYNE GACKENBACH
A Web Based Course on Dreams: Teaching Personally Sensitive Material Over the Internet

CAROLYN GALE, JOHN R. BOURNE, MARTINE DAWANT & DORIS QUINN
Implementing engineering education on the WWW: 3 case studies

PAULA GARRETT & STEVE RYAN
Reorganised by the Web: Bringing the Missing Link into Focus

GREG GARVEY
GENDERBENDER: Gender and the Human computer user Interface

LAURIE GELB
Health care information on the web: Model and menace

PHILIP GREENSPUN
Shame and War Revisited: Adding Semantic Markup to HTML

KEITH A. HALL & WILLIAM GATHERGOOD
Professional development for educators via the Web

DELWYN HARNISCH, VINCE GRACIA & MATT MARKUS
Web of thought: Ways to make the world web educationally relevant

AMIR HEKMATPOUR
On teaching and collaborative development of net-based multimedia knowledge based systems

MARK D. HENDRICKS
Distance education on demand: Serving MPEG on the Internet

MISOOK HEO, EMILE L. MORSE, SWANTJE WILLMS & MICHAEL B. SPRING
Multi-level Navigation of a Document Space

FREDERIC HOOGSTOEL & GREGORY BOURGUI
Using a WWW server to access and manage a Gemstone Smalltalk server supporting a virtual collaborative learning organisation

RICK HOROWITZ
Avoiding Vendor Lock-in: An Open Approach To Building Internet-Based Applications

IAN HOWATT & BASIA SIEDLECKI
Web Courses for Northern B.C.: Virtual Community, Virtual Pagination

ARTHUR C. HUNTLEY
Harvesting the Internet: Creating an on-line resource on the treatment of skin disease using the deliberations of an e-mail discussion group, RxDerm-L

PENNI LEIGH JONES
School participation in the co-development of commercial, Web-based curriculum resources

SOOK-HI KANG
Current situation of education utilizations of computer networks in Korea

CATHY C. KAUFMAN
The Web as an instrument in redesigning the teaching/learning transaction

CAROL KILPATRICK
Community based web projects to support human-computer interaction education

RICHARD KLEVANS & MLLADEN VOUK

Automatically generated web-based multimedia presentations

DAVID RIBEIRO LAMAS, JENNIFER JERRAMS-SMITH, & FELIZ RIBEIRO GOUVEIA
CAIN: Computer Aided Information Navigation: project description

IAIN LANGDON
Education for Changing Times - An Online Learning Framework

DMITRY LATUKHIN
Webbization in Russia

ELAINE LAWRENCE & ROBERT HUDSON
Creating a presence on the World Wide Web: Some key strategies

LISA M. LEHMAN & JOHN A. LEHMAN
Student reactions to a World Wide Web-Based Textbook: An Exploratory Study

FRANK LOCHTER & SVEN KNAUER
Development of 3W graphical information systems

CAROL LONG, SUE REEHM & MARY ITO DENNISON
Developing technological and content knowledge through publishing web pages

J.M. BRISSON LOPES & JOSE CARLOS TEIXEIRA
IDEALS: An approach to Web based remote training

RORY MCGREAL
A virtual campus on the World Wide Web

WILLIAM MEADOR
Information technology and the Viking Network

EUNICE M. MERIDETH & HILDA L. WILLIAMS
Start Your Search Engines! Authentic Activities and the Internet

CRAIG MONTGOMERIE & DWAYNE HARAPNUIK
Observations on web-based course development & delivery

CRAIG MONTGOMERIE & GARY DUGUAY
An alternative to brower-based instructions

CONWAY MOTHOBI
Using the World Wide Web to creatively support science education in schools

J. T. MUNK, A.T.M. AERTS & P. M. E. DE BRA
OODB support for WWW applications: Disclosing the internal structure of hyperdocuments

SHELIA NAUJOKS, LAURIE GRIFFITHS, & THEODOROS ARVANITIS
An Internet information service for the domain of personnel and development

GERMAN OTALORA & RAMSES HERRERA
A virtual classroom in the Web

VALERY A. PETRUSHIN & JANET KOLODNER
Put your experience on the Web: A tool for creating and browsing in case libraries

MARVIN L. PEYTON
The cyberquest project: A pilot project connecting K-16 Educational Institutions to Internet via High Speed Data Access

RICHARD PICKING
Incorporating world wide web technology into a multimedia learning environment

MARIA PIMENTEL & JOHN BUFORD
Non-invasive improvements to hyperlinking with HTML

JYRKI PULKKINEN & ESA NIEMI
WWW Technology and New Models of Learning

DAVID REIDER
World Band: K-12 Music Education on the Net, Collaborative Learning Environments

SERGE REZZONICO & DANIEL THALMANN
Browsing 3D bookmarks in BED

STEVEN RHEINTGEN
SEEWeb-Internet Solution: Novel application of the web

BERNARD ROBIN & SARA MCNEIL
Archive.edu: A storehouse of classroom resources on the Web

ED RODGERS
Description and evaluation of the Web-based course, 'Exploring the Internet,' at the University of West Florida

HENRYETTE ROTH & TIMOTHY CLARK
Electronic publishing and the Web

JAIME SANCHEZ & OMAR ALONSO
A Web-based tool for evaluating user-computer interaction

ROBERT SAPP
Providing real time instruction on the WWW: Lessons taught and learned

PAUL G. SHOTSBERGER & CHRISTOPHER G. SPELL
Professional Development On The World Wide Web: Training Secondary Mathematics Teachers

JULEAN A. SIMON & DAVID C. WOLHART
Discursive networking: context generalization in a networked environment

DAVE SNOWDON, LENNART FAHLEN, & MARTEN STENIOUS
WWW3D: A 3D Multi-User Web Browser

ELIZABETH SOMMERS
Convergent computer research at the undergraduate level: The World Wide Web as interdisciplinary teaching and learning tool

NANCY BUTLER SONGER
Kids interacting with kids world-wide through Web-based visualization technologies: The Kids as Global Scientists project

ANDY STONE
Project Pilot - The Application of Telematics in Co-operative Education

EUGENE S. TAKLE & MICHAEL R. TABER
Use of the Web as a tool for interactive learning

RUMI TATSUTA, KIMIKO GUNJI, MARY ELLEN MICHAEL
The Development of Multimedia Courseware of Japanese Culture via Internet

BONNIE THURBER & VITO DIPINTO
Middle school students as physical scientists: The practical and the virtual

A.A. TSINAKOS & K.G. MARGARITIS
Mentor: Internet search advisor and information retrieval system

YURY TSYBENKO
Feedback and adaptive interaction for WWW-based course

MAX A.TURNER & JAMES H. CAIRNS
Silicon soils: Teaching soil science via the Internet

JEAN VERMEL
The ETT Virtual Classroom: A Different Flavor

ANTONIO VOLPENTESTA, VINCENZO PISCULLI, ANTONIO GRECO, & NICOLA FREGA
"A Cantina i Bifarelli": A Java based environment for real-time multiplayer games.

ANA VON KLOPP
Interactive Teaching Material for the World Wide Web: A Syntax Tutor Written in Java

DAVID J. WANG & GEORGE CIERVO
Providing Enterprise Wide and Localized Searching at a Large, Decentralized Institution

DAVID WATSON & KATHY WATSON
Distant Education Materials via the Web - An Australian View

MATHIAS WILL
The virtual university: Making it happen

FAWN WHITTAKER
Timely tricks for Web language teaching

ITZIK YARHY, OFER MARGONINSKI, DAVID RASHTY & NAVA BEN TZVI
The Bible and Talmudic literature on the Web

JIA ZHUOSHENG & WEI HUOQING
The platform of corporate development and management: CAT software on the Web

PANELS

M J MCALISTER, P COMER & MISS K HADLEY

The Internet & The WWW: Friend or Foe?

TUTORIALS

ANDREAS BITTORF
How to set up and maintain an image database on the WWW efficiently

---

Copyright © 1996 AACE / Association for the Advancement of Computing in Education & The Web Society
This page created by Jerry D. Price, Jr. on September 23, 1996

Designing Hypertext Navigation Tools

Brian Bevirt

Digital Information Group, Scientific Computing Division,
National Center for Atmospheric Research, USA

Abstract: A common problem with hypertext information systems is that their organization is obscure, and this can hinder users. Developers can address this problem by providing navigation tools that help users find their way to the information they need. But before you can develop effective navigation tools, you must first understand the organizational scheme of your information system. This step--developing a simplified picture of your information system--can be the most difficult task in creating effective navigation tools. This paper describes some useful ways you can visualize the organizing principles of your information. Then it describes a variety of navigation tools that can reveal the structure of your information to your users: home page link, reference table links, "previous" and "next" links, table of contents, image maps (hierarchy charts, grids, other), search engines, and indexes.

Navigation Tools Can Keep Users Oriented

Navigation tools are features of hypertext documents that allow people who browse those documents to link to another location. Navigation tools like "previous" and "next" buttons are basic forms; more sophisticated tools like graphical maps also help *orient* the person who is browsing the information. The purpose of a navigation tool is to help people find only the information they need, so they can move on to their next task.

This paper describes an approach that will help you design and plan navigation tools to maximize the usability of the information in your hypertext documentation.

Standard Terms for This Presentation

To standardize the terminology in this presentation, I use the language appropriate for a World Wide Web site that provides HTML-coded files for people using browsers like Netscape. A "document" is a hypertext file that contains any combination of text, graphics, and audio. A "web site" is a collection of related documents provided by a hypertext server (by this definition, one server may have multiple web sites). A "browser" is a software application capable of displaying HTML-coded text and graphics on a computer screen. A "user" is the person who is using the browser to access the information.

Navigating Web Sites Can Be Confusing

Web sites can confuse users simply because there is no pattern for this new kind of "place" embedded in our cultural consciousness. Put another way, there seem to be no widely agreed-upon spatial topologies or other organizing principles for a multi-dimensional electronic information space [Lynch 1995].

When you hold and read a book, you have a wealth of visual and tactile cues about the information it offers you. Further, centuries of cultural tradition have standardized an extensive background of knowledge about how that information is presented: with only a few exceptions, a book has a title on the cover or the spine, it is organized in a linear sequence (starting at the front and ending at the back), it has page numbers, headings for chapters and sections, a table of contents near the front, and probably an appendix at the back that includes an index. And if you have a basic education, you know what all of these things are and what they mean to you as a reader.

A hypertext document in a web site is not a book; it does not come packaged with familiar cues and a long-standing cultural tradition. No organizing principles have yet been widely agreed upon for hypertext information systems. Yet many useful web sites exist. The following approach can help you design navigation tools to make your web site more usable.

Adapt Familiar Cues to the New Medium

Sounds easy, doesn't it? Well, it's easier said than done! But much of the groundwork has already been laid. You can use the information in this presentation to gain a perspective that will help you make this adaptation. I am not offering an automated tool to do this for you; you will have to think and plan and test and refine, but the result will be valuable to your users.

Many cues that have evolved with print media can be adapted to hypertext documents: title, heading, table of contents, index, glossary, diagram, and map. The challenge is to adapt them well, in a way that will be as intuitive as possible to your users.

Before proceeding, please recognize that effective organization of the information in the documents and of the documents in the web site is more important than the navigation tools. Just as good documentation cannot "save" a poorly designed and implemented product, good navigation tools cannot save a disorganized web site!

The site analysis I will recommend may help you improve the organization of your web site, but the purpose of this paper is to help you design good navigation tools. So here are two important characteristics of effective navigation tools.

Navigation Tools Must Be Predictable

In this paper, I assume that your web site exists to provide valuable information to your users. Navigation tools exist to help users find needed information and to keep them from becoming confused or disoriented while visiting your web site. Well-designed navigation tools are predictable: after a very small amount of experience with the tools, your users should understand how each tool will behave every time they use it. Each navigation tool must be consistent in how it is used and the result it produces, or it could confuse your users instead of helping them.

Multiple Navigation Tools Should Be Offered

Different people have different perceptual styles, and you can satisfy their differing needs by offering them multiple ways to access your information. Also, your users will have different levels of skill and experience, so their preferences for using your navigation tools will vary. You must first consider the size, content, and purpose of your web site to decide which navigation tools are appropriate. If your web site is large enough to benefit from *any* navigation tools, then you will help your users by providing more than one type.

Develop a Model of Your Web Site's Organization

This is the critical part of the work: you must develop a simplified but revealing model of the information in your web site, then you can prepare navigation tools that clarify its organization. (Please recall that I define a web site as a collection of *related* documents.)

An interconnected web of documents can express information and ideas with new depth and richness, but it also has the potential to confuse and intimidate users. An intimidated user will not receive full value from your site. Well-designed navigation tools will highlight your site's benefits to users and help them access the information they need.

The following steps are the most difficult part of the process. I recommend that you assemble a team of interested people with diverse viewpoints to participate in this process and help keep your creativity and insight fresh. Teamwork can lead to conflict, but conflict can also generate the fire that forges useful new ideas.

Identify a Single Sequence for Your Documents

Your first attempt to simplify your web site should be to define a linear sequence for every document in it. Review the organizing principles you had in mind when you created the site, and think about how you would arrange a table of contents for it. If you didn't create your site, or if it evolved in some less-than-logical fashion, then analyze the files in your site and determine what organizing principles can improve its usefulness to its users. Determine which documents provide overview information, which provide supporting information, and which provide results or concluding information. Then make numerous attempts to establish a sequence of documents within each of those three groups.

Examples of sequential structures that go beyond a plain sequence include sequences with alternate paths, side notes, shortcuts, backtracking, and loops.

Organize the Contents of Your Site into a Table or Grid

If you can't reduce the contents of your site to a simple sequence, try to map it onto a grid of rows and columns. The table is a familiar organizing structure in Western culture. A command reference manual--UNIX man pages for example--map conveniently onto a grid: each command has sections titled Name, Synopsis, Description, Options, and so on.

Define the Contents of Your Site as a Hierarchy

Your site may be conceptualized as a hierarchy. The hierarchy is a common and powerful way of organizing information and people. Research has shown that users prefer a hierarchical to a linear organization of menus. They make fewer errors in a hierarchical structure, and hierarchical structures result in clearer mental maps. Finally, technical users especially prefer hierarchical organizations [Horton 1994]. Examples of hierarchy diagrams include tree diagrams, flow charts, and the star configuration.

Create a Picture of Your Site as a Web

If all of the preceding ways to simplify the organization of your site escape you, create a picture of it as a web. To start, cut up some sheets of colored paper into rectangular cards. Each card represents a document. Write the file names of the documents on the cards. Use the different colors of paper to organize the documents into related groups.

Use string to represent the links between your documents and connect the cards. Don't try to represent *all* of the links between your files, just indicate the most important relationships. Determine which files belong together, then decide which ones make sense at the beginning of a group and which ones make sense at the end of a group. Determine how the groups relate to each other.

If you still can't make sense of your site's organization, your users have no chance!

After graphically representing your site as a web, start mapping it onto the simpler structures. Now can you draw a picture of it as a hierarchy, or a grid, or a sequence?

A diagram of a web is unlikely to help users recognize and remember the organization of your site; I only recommend that you design navigation tools based on the simpler models. You can also use the colored-card strategy to structure your documents according to any of the other models.

Break Through to a Fresh Viewpoint

If you have seriously attempted to create a simple model of your site but are still having trouble, it's time for you to refresh yourself by taking a break. Consider this analogy: if you shine a bright light on a three-dimensional object, it casts a two-dimensional shadow. The shadow is a simplified projection of the physical object: it is just an outline. A simple model of your web site is also an outline!

Let's say our three-dimensional object is an ordinary pair of scissors. Now consider this: if you close the scissors and point its tip directly at the light, the shadow will not reveal much information about the shape or the size of the scissors. In fact, most people who see only the shadow probably won't even recognize what kind of object is casting it. But if you open the scissors and shine the light at it broadside, its shadow will be immediately recognizable to almost any observer.

So bear in mind that some simplified models of your site will be more revealing than others. You can dramatically improve a model just by using a different perspective. This is how a team of people with diverse viewpoints can help you model your website. Some models should include more detail than others. And the individual differences between users will give them contrasting preferences, so it is best to develop more than one model.

Create a Suite of Navigation Tools from Your Models

The difficult work is now complete. All you need to do now is select the models that work best for your site and your users, render those models as text and/or graphics, then install the links in the navigation tools, and install the links *to* the navigation tools from every document in your web site.

The links to the navigation tools should be clearly labeled and consistently located. While some icons are likely to be recognized by most people, most graphical symbols that stand for abstract concepts are ambiguous, unclear, and even misleading. You may think that users can learn what your icons represent, but you shouldn't depend on it: always label your icons with helpful text.

If all of your documents are very brief (about one screenful), then you can place the links to your navigation tools at the beginning or end (consistently one place or the other!) of each document. If many of your documents are two or more screenfuls in length, then I recommend that you place links to your navigation tools at both the beginning and end of each document.

Link to the Home Page and Reference Pages of Your Web Site

It is customary to offer a link to the first document (home page) of your web site as a navigation aid. Users expect to be able to orient themselves by returning to this top-level document because good home pages typically provide useful information about and links into the web site.

If your web site provides other documents dominated by useful links--reference tables for example--you can help users by offering standard links to those documents along with your other navigation tools.

When To Use "Previous" and "Next" Links

Many web sites are easy to organize into a linear sequence because that is how they were conceived and implemented. If your site provides information in a well-defined order, and if this order is necessary for users to obtain full benefit from your site, then a "previous" link at the beginning of the document and a "next" link at the end of the document is appropriate.

However, just because you can define a sequence for all the documents in your site does not mean that you should include previous and next links. These links imply to users that the linear sequence is important for understanding the material. You should include them only if this is true.

A Table of Contents is Highly Recommended

A table of contents is an almost universally understood navigation tool. It outlines any body of work by listing document titles, headings, and sometimes subheadings. Low-level subheadings are normally not included in the table of contents because they can expand the table of contents to an unwieldy length.

Many users who are new to hypertext documentation feel welcome when they find a web site's table of contents because it is comfortable and familiar to them. It also helps introduce them to the benefits of hypertext links because they no longer have to look for page numbers to find the information they need.

The linear-sequence model of your web site provides your table of contents.

You can add levels of meaning to your table of contents by indenting subordinate or supporting titles and headings under the more general titles and headings. An indented table of contents maps perfectly to a hierarchy chart, and vice versa. That is, if you can organize your documents into an indented table of contents or a hierarchy chart, then you can *and should!* provide both for your users.

Hierarchy Charts Offer Numerous Benefits

Hierarchy charts may differ from tables of contents in the amount of links they provide. Hierarchy charts can provide a spatially organized overview of an entire web site: users can see the structure of the site in a single overview. And through image mapping, you can let users move directly to any document in the web site by pointing and clicking on the titles in the chart. But for larger web sites, there is no room for long titles or too many entries on the chart.

So titles will need to be shortened for use on the chart. And more headings will probably need to be omitted. But even if multiple hierarchy charts have to be created to cover different parts of the web site, user benefits from well-designed charts used as navigation tools are worth the effort it takes to create them. Examples of multiple hierarchy charts as navigation tools appear as the "roadmaps" in the 1995 NCAR Annual Scientific Report at http://www.ncar.ucar.edu/archives/asr/ASR95/ASR95home.html [NCAR staff 1996].

You can add value to a clickable chart by highlighting the title of the last document the user visited before coming to the chart. This "where-am-I" feature provides additional orientation cues for users. An example of this feature appears throughout the NCAR Graphics interactive documentation at http://ngwww.ucar.edu/ngdoc/ng4.0.1/nggenrl/ngroadpg.html [NCAR Graphics Group 1996].

Grid Charts Can Add Value

If the organization of your site resolves to a grid or a table, you can create a grid image map to allow users to navigate by pointing and clicking on the titles in the grid. Grid charts provide many of the same advantages as hierarchy charts.

Other Kinds of Images Must Be Simple

Image mapping offers a wide range of possibilities for creative navigation tools. Unfortunately, it also offers an equally wide range of possibilities for confusion. The image must make clear which areas are linked to other files and which areas are not. The linked areas must also clearly indicate where the link leads; this can be extremely difficult unless you know that everyone in your audience thinks just like the image designer. Factors such as cultural conditioning, color blindness, and different perceptual styles can frustrate the most valiant attempts to develop an "intuitive" image. An example of an intuitive image map is a geographical map that shows political boundaries; the area within each boundary is mapped to information about that political entity.

Search Engines Need Introductions

Many sites are providing a search engine to help users find information in their site. While this idea is admirable, the search engine must be implemented in a user-friendly way. Many sites provide insufficient introduction to the purpose and content of the site. Some sites even greet users with a statement like "This site is searchable. Type a keyword in the box." The user never received a proper introduction and is now facing an empty box.

Many different search strategies are arising in response to the evolution of search engine technology. So the user introduction should also provide concise, helpful information about how to get results from your search engine. Linda Barlow of Monash Information Services publishes an excellent and informative resource to help you understand search engines at http://www.monash.com/spidap.html [Barlow 1996].

Indexes Have a Long Way to Go

Hypertext indexes are still rare. Some are nothing more than an alphabetized collection of all the links at the web site. Users who are very patient or very desperate might use such an index, but this is not a user-friendly navigation tool. A well-developed index--one that simulates what you would see in a high-quality printed book--can actually help users learn about the information at a web site. The index entries and the entries indented below them can provide useful insights about the content of the web site, especially if a variety of keywords have been indexed for each of the most important concepts. But hypertext indexes designed like print-media indexes are extremely rare. This is probably because they are difficult to develop and and even harder to maintain as web sites evolve. But such indexes would be a welcome addition to a suite of effective navigation tools.

References

Lynch, P. J. (1995). Web Style Manual. New Haven, CT: Yale Center for Advanced Instructional Media. http://info.med.yale.edu/caim/stylemanual/M_I_2A.HTML

Horton, W. K. (1994). Designing and Writing Online Documentation, 2nd Edition. New York: Wiley. http://www.horton.com/brochure/dwold.htm

NCAR staff. (1996). 1995 NCAR Annual Scientific Report. Boulder, CO: National Center for Atmospheric Research. http://www.ncar.ucar.edu/archives/asr/ASR95/ASR95home.html

NCAR Graphics Group. (1996). NCAR Graphics Version 4 Interactive Documentation. Boulder, CO: National Center for Atmospheric Research. http://ngwww.ucar.edu/ngdoc/ng4.0.1

Barlow, L. (1996). The Spider's Apprentice, How To Use Web Search Engines. New York: Monash Information Services. http://www.monash.com/spidap.html

Developing Structured WWW-Sites with W3DT

Martin Bichler, Stefan Nusser
Martin.Bichler@wu-wien.ac.at,
Stefan.Nusser@wu-wien.ac.at
Vienna University of Business Administration and Economics
Department for Management Information Systems
Augasse 2-6
1090 Wien
Austria

Abstract: In this paper we introduce W3DT (World Wide Web Design Technique), a semi-formal design technique which is capable of modeling the dichotomy of structured and unstructured information represented in large Web-sites. Hence, it is suitable for modeling the WWW-frontend of a database system as well as for designing the hypertext description of a company. The proposed graphical notation provides an easily comprehensible framework for the development and rapid prototyping of a WWW-based information system. The design process is further facilitated by the W3DT Web-Designer, a prototypical application, which supports the graphical notation of the methodology and is capable of generating a framework of HTML-pages and CGI-Scripts at every step of the design process.

Introduction and Motivation

For a lot of organisations the WWW provides a new and powerful opportunity to interact with customers. The scope of these sites covers the presentation of products, the company's business performance as well as the announcement of job offers. The extent has grown due to the numerous tasks they fulfill. A few hundred pages and various gateway scripts to inhouse information systems have to be planned and implemented. At the moment we lack guidelines and tools for a structured design process of such hypermedia applications and therefore the ever growing network of interlinked applications is becoming increasingly confused and hard to maintain [see Bieber et al. 1995].

In conventional software engineering, design and development methodologies have been established to cope with complex tasks. Some of these experiences can also be adopted for the development of Web-sites. But Web projects differ from traditional software development projects in several critical dimensions. They involve people with very different skill sets: authors, content designers, artists as well as programmers. The design of hypermedia applications involves capturing and organizing the structure of a complex domain and making it clear and accessible to users. After all, the multimedia aspects of Web-sites raise numerous difficulties. Therefore, [Isakowitz et al. 1995] emphasize that hypermedia design is a challenging process that is currently more of an art than a science. Nevertheless, a solid development cycle and formal design techniques can help developers substantially. They lead to better Web-sites, which means better to read, to navigate and to maintain.

W3DT (World Wide Web Design Technique) was invented especially for the development of large Web-sites. The W3DT methodology is capable of modeling highly structured, database-like information-domains on the one hand and conventional hypertext on the other and thus represents the dichotomy of structured as well as unstructured information. A prototypical tool called Web-Designer supports the graphical notation of the methodology and is capable of generating a framework of HTML-pages and CGI-Scripts at every step of the design process.

This paper is structured as follows: In the next section, the current approaches to hypermedia design are described and the need for new concepts is explained. We refine these needs in the following section and briefly show the basic W3DT design primitives and their use. The main section gives an example application which is modelled and generated step-by-step using the W3DT Web-Designer.

W3DT and other Approaches to Hypermedia Application Design

Hypermedia design has to consider many different aspects as for example information structuring, navigational design and interface design. Database design is a field very closely related to hypermedia design. Many formal hypermedia design methodologies are based on well defined database design methods like the Entity-Relationship model. In database design, models play a crucial role to express the intrinsic application oriented data semantics. Some of the most popular hypertext design methodologies at present, like RMM, HDM or it's successors HDM2 or OOHDM adopt modeling constructs of database design and add hypermedia related features like navigational constructs see [Isakowitz et al. 1995], [Garzotto et al. 1993].This approach has been chosen because it is familiar to system analysts. This is the main reason, why these methodologies are useful for applications with highly structured information and high volatility (DBMS interface, product catalog).

On the other hand, practitioners tend to draw the nodes and links of their hypermedia applications in a relatively straightforward manner in order to convert linear text into hypertext structures. The Web supports unstructured information like multimedia news services as well as interfaces to databases or other kinds of structured information. Thus a design method for WWW-based information systems requires representation of both. W3DT provides the necessary modeling constructs to fill this gap. The notation is intentionally kept as simple as possible to make W3DT easily comprehensible for the novice user.

Information Structuring and Navigation Design with W3DT

Since W3DT is designed for modeling WWW-sites, a *Site* is the broadest concept in W3DT. Each *Site* consists of at least one *Page* and optional *Links*. Figure 1 shows the graphical representation of the elements in the W3DT modeling technique and gives a short definition of each element.

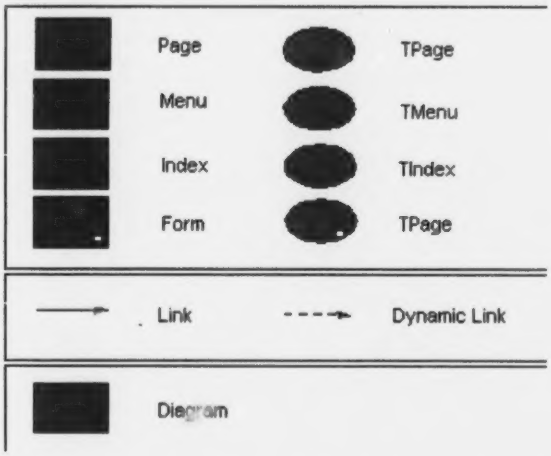


Figure 1: Elements of the W3DT design methodology

Page, Index, Form and Menu are the basic means for structuring the information domain of a hypermedia application. Note the difference between an Index and a Menu - the former is used to model a complete enumeration of links, like a complete list of links to the members of a faculty. The latter is a navigational aid consisting of links to other pages. An example is the homepage of a site. Its main purpose is not to give a complete list of all its pages, but to give the user navigational aid and direct access to the main topics of the site.

Each of these design primitives can also be used as a Template (TPage, TForm, TIndex and TMenu). The Template-notation is very similar to the concept of an Entity-Type in the Entity-Relationship model. The pages created by a gateway-script querying a database are modeled by this construct. For example a TPage named Faculty-Member can be used to denote similar descriptions of all the faculty-members.

Usually, pages generated by one gateway-script on the fly have the same layout, but the information displayed differs according to the script's input arguments.

Links and Dynamic Links are used to model the Navigational Design of an application. Hardcoded links are the standard way to navigate between WWW-pages. The Dynamic Link is slightly different - it represents the execution of a gateway-script. All this leads to a rather simple rule: Links refer to static objects, Dynamic Links refer to Templates. From the user's point of view, there is no difference between a Dynamic Link and a conventional Link; a Web-Page generated on the fly looks exactly like a hardcoded page. For the analyst, the difference is real - a Dynamic Link has additional attributes (the path and implementation language of the gateway-script, it's command-line-arguments etc.) which she has to take care of.

The last element shown in Figure 1, the *Diagram,* is used to denote a hierarchical refinement. The only purpose of this construct is to reduce the number of elements in a diagram and make it more comprehensible.

Since the W3DT method was developed for the task of designing a WWW-site, it is best illustrated by regarding the construction of a small WWW-based information-system.

Example: The Cybersound Web-Site

Cybersound is a record shop which intends to create a WWW-site. The existing RDBMS shall be used to provide up-to-date information on products and prices. This feature shall be implemented with a search-engine, giving the user the possibility to fill out a form and get a list of matching disks with their prices. Furthermore, general information on the company shall be displayed as well as information on Cybersound's staff and a list of job offers. The description of Cybersound's business activities is provided by the marketing department and not expected to change very often. The information on Cybersound's employees comes from another database in the personnel department. Since this data changes frequently, the corresponding Web-pages should be generated on the fly. Finally, as a special customer service, Cybersound wants to maintain information on a number of cultural events. Figure 2 is a screenshot of the W3DT-Web-Designer showing a model which represents this information.

The navigational design in this simplified example is very easy. The homepage is a Menu - a page with the main purpose to present and explain the meaning of the parting hyperlinks. From Cybersound's homepage, the user can move on to one of the three main areas.

CD-Form is the starting point for the WWW-based interface to the product database. This page contains a form with the input fields necessary to restrict the number of records found. From this Form, the user moves on to a TPage, a dynamically generated Page, which will present the matching records. The page Records is modeled as a Template since the information presented to the user (the data fetched from the database) is retrieved on the fly. The link from CD-Form to Records is a Dynamic Link. As a rule, every link leading to a Template must be a Dynamic Link.

The general information on Cybersound is structured by the Menu *Company*, which presents links to Cybersound's business activities as well as to the Index *Staff*. Note that the information represented in the Pages *Business Activities* and *Hiring* is of static nature - whenever the content needs to be updated, the corresponding HTML-File needs to be modified.

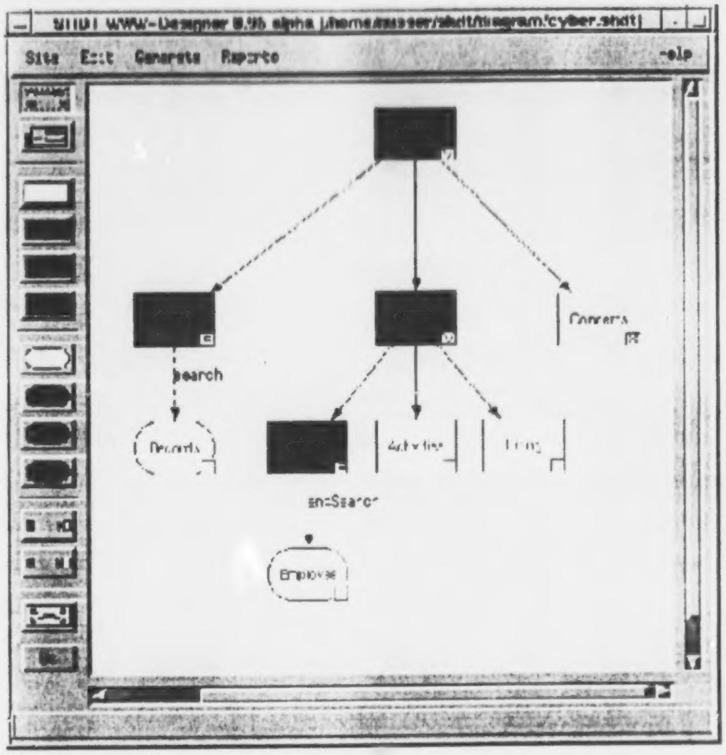


Figure 2: The W3DT Web-Designer Showing a Model of the Cybersound Site

The Template *Employee* is linked to the Index *Staff* by means of a Dynamic Link. The meaning of this construct is straightforward: The Index contains a static list of all employees (whenever a new employee is hired, this list has to be updated manually). The user can choose one of the names presented in this list and will get further information on the selected person from the database. This dynamically generated Page is modeled by the Template *Employee*.

Finally, the model shown in figure 2 contains the Diagram *Concerts*, referring to a submodel which describes this domain in more detail.

This small example gives an impression of the W3DT modeling constructs. The W3DT modeling technique is described in more detail in [Bichler and Nusser 1996]. In the next section we will show how the Web-Designer can be used to generate a running prototype of the specified information system.

Computer Aided Web Development with the W3DT Web-Designer

There are several ways to support a developer in the design phase. Of course a formal design methodology is an important part. But to push efficiency in this development step, a computer aided design tool for WWW-based information systems is needed. Figure 2 shows a screen shot of the W3DT Web-Designer, a CASE-tool for the WWW. In developing the Web-Designer we had the following goals in mind:

- A tool should completely support the graphical notation of the design methodology.
- The definition of Layouts (comparable to the style-sheets of HTML 3.0) should facilitate the design of uniformly looking Pages throughout the whole site. Although the pages generated by the Web-Designer are only a skeleton which is completed by means on an HTML-editor, this feature has proven to be very useful if a large number of files are involved.
- The user should be able to generate a running system at any time throughout the design process.

Thus we have fast feedback loops on each design decision.

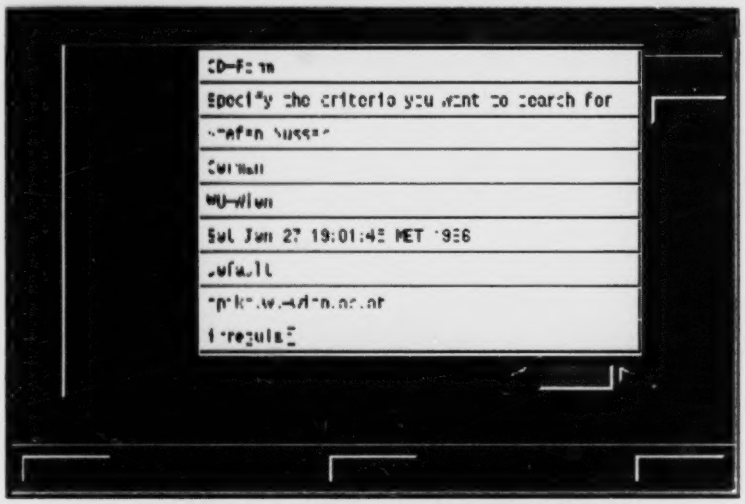
- Using the HTML-pages and CGI-scripts generated by such a CASE-tool, developers should have the possibility to better communicate their ideas to other people. In short we intended to provide a powerful prototyping tool, which on the one hand saves development time and on the other is a helpful means to meet the needs and wishes of a customer.

The W3DT Web-Designer was developed using Perl 5.0 and Wafe [see Neumann and Nusser 1993], a prototyping environment for the X-environment. The Web-Designer provides an easy to use graphical browser for the W3DT methodology. The user is able to draw and edit all the W3DT-constructs in a very straight-forward manner. Large sites with several hundreds of pages are modeled within very short time. By means of the hierarchical refinement, the models are kept clear and easy to comprehend.

Generating a prototype of the Cybersound site

In order to demonstrate the functionality of the W3DT Web-Designer, we will now take a closer look at the CD-Form part of the Cybersound example introduced in the previous section. The first step in generating a running system for this model with the W3DT Web-Designer is the definition of a Site. Essentially, this means that the user provides basic information on the target system. Among others, the directories for HTML files or CGI-scripts have to be specified.

Once a site is defined, the user can start to add the objects of the W3DT design methodology and to link them together with static and dynamic links as shown in figure 2. The button bar on the left side of the window provides easy access to all the necessary W3DT constructs.



CD-Form

Specify the criteria you want to search for

WU-Wien

default

Figure 3: General Form Attributes

Any time during the design process, new Layouts can be created. A Layout basically consists of a header, a footer and other attributes, which will be used in all pages and CGI-scripts throughout the system. Once a Layout has been created, it can be assigned to any page in the diagram. This helps to achieve a uniform "look and feel" of the Web-site. After having designed the information and access structures of the site it is necessary to set the page-specific attributes. Through context-sensitive windows all attributes of the currently selected element can be modified. Figure 3 shows as an example the attribute window of a Form.

Here all relevant information of the page like its name, a comment, its language, or the responsible organization can be entered. Much of this data is used for the generation of helpful reports for the administration of a site. By means of the index-tab *Form* all dialog elements of a Form can be defined (shown in figure 4). Text fields, check boxes, selections and all the other form elements supported by

BEST COPY AVAILABLE

HTML 2.0 can be inserted. Layout and form elements are important attributes for the generation of a prototype of the designed site.

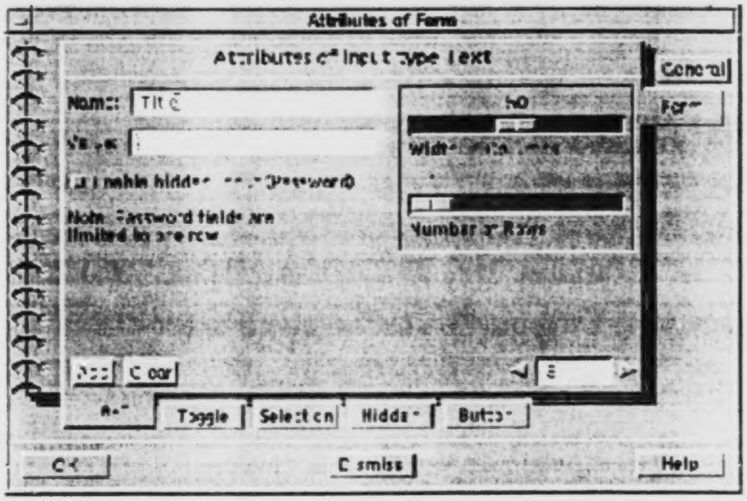

Attributes of Form
Attributes of Input type Text
General
Form
Name:
Value:
Width
Number of Rows
Toggle
Selection
Hidden
Button
OK
Dismiss
Help

Figure 4: Input Arguments of a Form

Any time during the design process the Web-Designer is able to generate running prototypes of your Web-site. The specified layouts, page attributes or form elements are converted into HTML and Dynamic Links are translated into CGI-scripts producing the specified Templates (TPage, TIndex, TMenu, TForm) on the fly. The input-parameters of the script are the form elements or the index of the outgoing page.

The whole process of generating this part of the Cybersound example is illustrated by Figure 5.

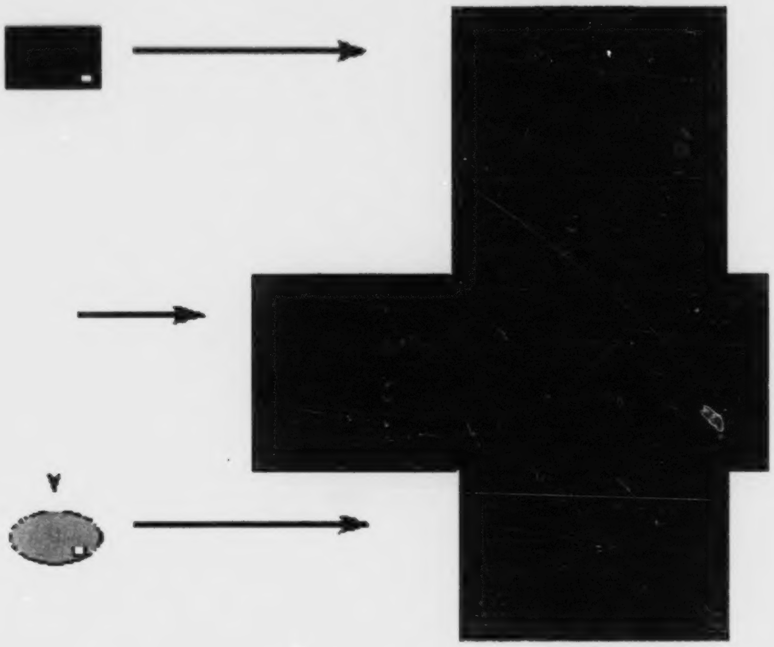


Figure 5: W3DT Generation - Overview

BEST COPY AVAILABLE

Conclusions

It was shown that the design of large Web-sites has special needs requiring new concepts. The W3DT project provides some powerful new ideas. It is an approach to support developers as far as possible in designing and prototyping large Web-sites. Many existing Web-sites have been modeled successfully. We are currently testing the W3DT Web-Designer with several groups of students working together with companies in order to design and implement commercial Web-sites. We hope that W3DT helps developers of large Web-sites to a more efficient and more structured way of doing their job.

References

[Anderson 1995] Anderson, C.: The accidental superhighway, in: *The Economist*, July 1st 1995.

[Bichler and Nusser 1996] Bichler, M.; Nusser, S.: W3DT - The Structured Way of Developing WWW-Sites, *Proceedings of the ECIS 1996*.

[Bieber and Isakowitz 1995] Bieber, M.; Isakowitz, T.: Designing Hypermedia Application, in: *Communications of the ACM*, Volume 38, S. 26 - 29.

[Garzotto et al. 1993] Garzotto, F.; Paolini, P.; Schwabe, D.: HDM - A Model-Based Approach to Hypertext Application Design, in: *acm Transactions on Information Systems*, January 1993 - Volume 11 No. 1.

[Garzotto et al. 1995] Garzotto, F.; Mainetti, L.; Paolini, P.: Hypermedia Design, Analysis and Evaluation Issues, in: *Communications of the ACM*, August 1995 - Volume 38, S. 74 -85.

[Gloor et al 1990] Gloor, P. A.; Streitz, N. A.: Hypertext und Hypermedia. Informatik-Fachberichte 249. Springer-Verlag, Berlin 1990.

[Halasz and Schwartz 1994] Halasz, F.; Schwartz, M.: The Dexter Hypertext Reference Model, in: *Communications of the ACM*, February 1994 - Volume 37, S. 30 - 39.

[Hansen 1994] Hansen, H. R.: Implementing Mass Information Systems - A Conceptual Framework and Guidelines. CITM Working Paper Series, No. CITM-94-WP-1004, Walter A. Haas School of Business, University of California, Berkeley 1994.

[Hofmann 1995] Hofmann, M.: Problemlösung Hypertext, Hanser 1995.

[Isakowitz et al. 1995] Isakowitz, T.; Stohr, E. A.; Balasubramanian, P.: RMM: A Methodology for Structured Hypermedia Design, in: *Communications of the ACM*, August 1995 - Volume 38, S. 34 - 43.

[Nanard and Nanard 1995] Nanard, J.; Nanard, M.: Hypertext Design Environments and the Hypertext Design Process, in: *Communications of the ACM*, August 1995 - Volume 38, S. 49 - 56.

[Neumann and Nusser 1993] Neumann, G.: Nusser, S.: Wafe - An X-Toolkit based frontend for application programs in various programming languages, in: *Proceedings of the Winter 1993 USENIX conference*.

[Taylor 1995] Taylor, D.: Business Engineering with Object Technology, Wiley & Sons, 1995.

[Thüring et al 1995] Thüring, M.; Hannemann, J.; Haake, J. M.: Hypermedia and Cognition - Designing for Comprehension, in: *Communications of the ACM*, August 1995 - Volume 38, S. 57 - 66.

Developing an internet Section of a Management Course: Transporting Learning Premises Across Media[*]

March, 1995

John D. Bigelow
Professor, Management Department
Boise State University
College of Business and Economics
1910 University Drive
Boise, Idaho 83725

RMGBigel@cobfac.idbsu.edu
(208)385-1267

Suggested Conference Topic Area: The Web as Teaching Tool

Abstract

In universities we're seeing a rapidly growing interest in developing internet implementations of classroom courses. In translating courses from a classroom environment to an internet environment, how can we both build on the course development that occurred in the classroom environment, and avoid inadvertently limiting the possibilities of the internet course? The author proposes a "premise-based" approach, in which learning premises underlying classroom courses are extracted, then used as the basis for creating an internet course. This approach is applied in developing an internet section of an introductory management course. Six learning premises are identified and organized into a learning model. The author then describes how each element of the model is implemented in the internet course. Illustrations of supporting email, internet, and Toolbook resources are provided, and student reactions to the course are reviewed.

Contents

premises about learning	learning model	course goals	course structure	learning resources	learning activities	assessable outcomes	feedback	student response

---

Developing an internet Section of a Management Course: Transporting Learning Premises Across Media[*]

March, 1995

In the past couple of years we have seen a vast surge toward use of internet. Whereas a few years ago the internet was an obscure reserve for academics and researchers, we now are seeing in excess of seven million visitors a day. Moreover, increasing access to internet, coupled with decreasing educational budgets has led to interest in the internet as a teaching/learning medium. At the university level, internet-based course development has moved beyond the occasional experimental course, to more broad-based course development; e.g., through the Globewide Network Academy and The Online College Classroom. Because of the low capital investments required to establish internet courses and because of the vastly extended prospective clientele for such courses, we are likely to see a lot more of them in the future. The University of Phoenix, for example, is advertising its on-line campus in Business Week.

The potential of the internet as a university teaching medium, however, has still to be explored. To be sure, we have a teaching model for on-line learning which has been developed over years of BBS (bulletin board system) and some internet development. Yet this model was developed under fairly severe computing and communication speed constraints. We now have capabilities and resources available to online teachers which were scarcely dreamed of in the early BBS days. When developing an internet course then, it is important to keep in mind the possibilities which the contemporary internet environment makes available, and not to simply adopt designs which were tailored to past constraints. It is important to remember too that much of university experience in learning has evolved in a classroom environment. How can we take this learning process and creatively transport it to an internet environment?

One way of doing this is to extract the learning premises which underly classroom designs, and to then reconstruct a learning design based on these premises in an internet environment. By building an internet learning design based on established learning premises, an internet teacher is required to stay close to learning issues, and use these premises as a basis for choosing appropriate internet learning elements.

There are undoubtedly other ways of developing an internet couse; e.g., by adopting designs of other internet courses or by applying a theory of learning. The premise-transport approach, however, has a particular benefit in this circumstance, since the internet course is to be offered as one section of a multisection course. Since the course is a core course in our college, it is important to maintain a fair amount of consistency among what is learned among sections. The approach taken here promises to provide that consistency while at the same time allowing creative reconstruction of classroom premises in an internet environment. In addition by building the course on a stated set of premises, this approach enables other instructors to critically examine the underpinnings of this course and to modify them as needed when adopting the resultant design to the learning needs of other courses.

The purpose of this article is to describe such a premise-based process undergone by the author in developing an introductory management internet course. This article begins with a discussion of premises about learning which underly current learning structures. These premises are then applied so as to recreate the course in an internet environment. Finally student response to the course are reviewed.

Premises about Learning

One historical image of learning is of the schoolhouse, where children are chanting sums together. A teacher who organized this kind of activity might provide a number of explanations as to why children were doing this, and some of them would involve premises about learning; e.g., repetition is necessary for mastery, and this kind of effort builds self-discipline. Similarly, in college learning we see students engaging in activities, and again teachers would rationalize these activities in part using premises about how learning is accomplished and the kind of learning that is desirable. The author investigated learning premises underlying classroom sections of the introductory management course, both through conversations with teachers of this course and through assessment of his own experience as an instructor of this course. Six premises were surfaced concerning learning in this course:

1. **Learning as goal-based**. The development of a clear learning goal helps in developing a course designed to accomplish this goal. It helps students understand the course and to contribute to learning. It enables student and instructor to evaluate learning and to develop improved learning designs.
2. **Learning as resource-supported activities**. Learning occurs through activity. These activities often require learning resources, such as readings, communication channels to others, case information, etc. A course structure, then, can be thought of as a system of activities directed toward a learning goal, and supported by learning resources.
3. **Conceptual thinking**. At the heart of this course is concept mastery. This is accomplished in phases, involving study and understanding of concepts, followed by application practice. The latter may start with straightforward application of a concept to a situation. Later application may involve both the diagnosis of situations and application of one or more concepts.
4. **Use of feedback**. Timely and effective feedback to students about the quality of their work is necessary for student, instructor, and course learning. Feedback derives from assessable outcomes of activities.
5. **Active learning**. A "metagoal" of the course is to support students' abilities to learn proactively.

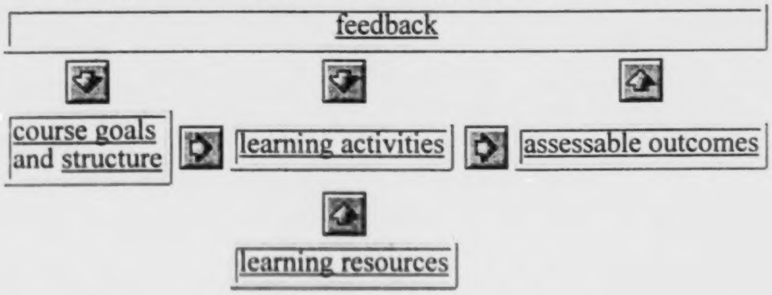
This is done through participation by students not only in target course learning but also in development of the learning structure.

6. **Learning motivation**. Student learning is a function of motivation. Extrinsic motivation for learning can be enhanced through a specific, challenging, and fair grading system, as well as instructor/class recognition of performance. This system, however, should be balanced with a program for developing intrinsic motivation for learning. Intrinsic learning is enhanced by student participation in the learning process, by increasing the perceived relevance of course learning, by developing a feeling of ownership and influence over the course, and by supporting individuation of learning so as to follow personal interest.

Other premises concerning the course were also surfaced, but did not have directly to do with learning. For example, the premise "if the course weren't so big, I would learn everyone's names and require more individual work" concerns activities felt to be potentially valuable by the instructor, but not feasible given the typical sizes of sections. Such premises could have relevance for an internet section, if it turns out that these constraints are lessened or eliminated in the internet section.

A Learning Activity Model

The above premises suggest a learning process containing the following steps:



The task now is to implement this set of premises in an internet environment. Figure 1: Steps of Learning Process

Application of Learning Premises to an Internet Environment

Course Goals

The introductory management course was chosen for internet delivery because: (1) the content of the course is primarily conceptual, and thus a more suitable prospect for an internet medium, (2) it is a multisection course, providing a choice for a large number of students, and (3) students with business majors may find this section distinctly valuable, either because it would help in developing a schedule, or because they are working and find attending any scheduled classes difficult. These assumptions will be revisited in the last section of this article.

The internet section of this course, then, was intended to accomplish the same goals as had been established for other (classroom) sections: *To provide a broad understanding of organizations as a basis for dealing with complex organizational issues*. Two objectives elaborating on this goal were also adopted:

1. *Participants' understanding of organizations will be expanded to include organizational:*
    - ***Management and Problem-Solving**: management and the process of addressing complex organizational problems.*
    - ***Environment**: the major elements of the larger context of organizations, including global competition, major technological changes, legal/political, and social issues.*
    - ***Adaptation**: organizations as structures designed to be adaptive to these elements, and different organizational and inter-organizational designs as means of doing so.*

BEST COPY AVAILABLE

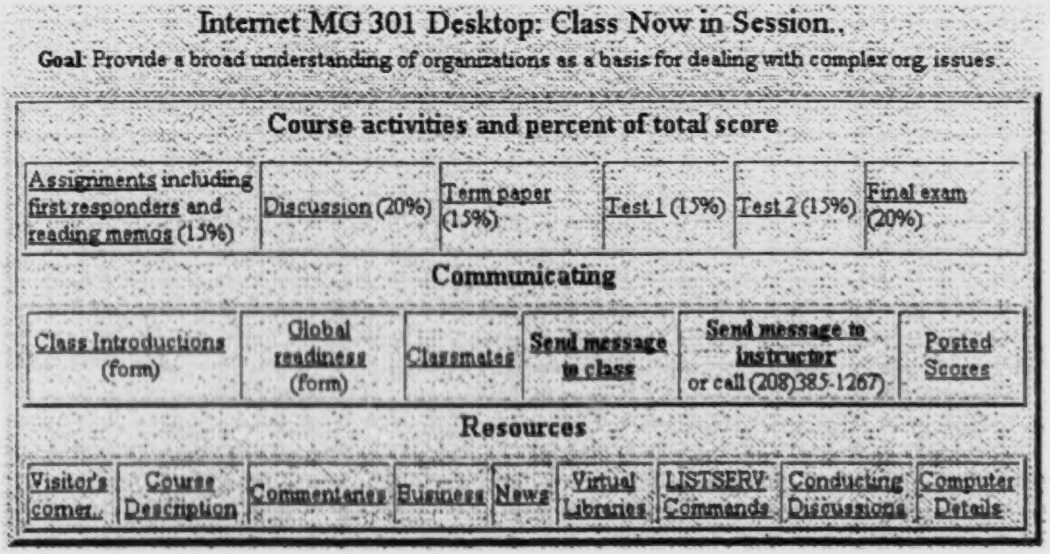
- ***Policies and Practices**: how organizational policies and practices (including culture, goal setting, diversity, leadership, motivation, and management information systems) shape or constrain goal attainment.*
- ***Innovation and Change**: ways by which firms may organize for change, and means of furthering the innovative and entrepreneurial aspect of contemporary corporations.*

2. *Given complex organizational situations, participants will be able to draw on the above to raise and analyze issues, create alternatives, develop solutions, and plan implementation.*

Course Structure

An initial concern was how to communicate to students the nature of this internet course, and to do it via an internet medium. The "tree branching" structure typical of internet sites is often difficult to understand, since much of the substance of the site is one or more levels away from the initial view. Exploration of branches can, unless the designer is careful, get the inquirer lost in a maze of interconnected branching links and make it difficult to get a sense of the course as a whole.

A three-rowed "desktop" was used as the primary entry point for the course: the elements of the course were laid out in a rectangular set of cells each linking to some aspect of the course. The "mg301 desktop" is shown in Figure 2:



Internet MG 301 Desktop: Class Now in Session..

Goal: Provide a broad understanding of organizations as a basis for dealing with complex org. issues.

Course activities and percent of total score

Assignments including first responders and reading memos (15%)	Discussion (20%)	Term paper (15%)	Test 1 (15%)	Test 2 (15%)	Final exam (20%)

Communicating

Class Introductions (form)	Global readiness (form)	Classmates	Send message to class	Send message to instructor or call (208)385-1267	Posted Scores

Resources

Visitor's corner..	Course Description	Commentaries	Business	News	Virtual Libraries	LISTSERV Commands	Conducting Discussions	Computer Details

Figure 2: MG 301 Desktop

This desktop contained three rows. Since, in the author's experience, students wanted first to know about class activities and scoring, these were arrayed along the first row of the desktop. The second row contained communication tools, allowing students to communicate with the instructor and one another. The third row contained information resources, including links to internet, comments from the instructor, and help on computer and discussion aspects of the course. The "visitor's corner" was aimed toward prospective students, who may preview the site while the course is in progress. By arraying elements of the course in correspondence with the way students are likely to think about a course, this desktop design attempted to make the course as easily comprehensible to students as possible.

Learning Resources

The types of learning resources available to students in an internet course are different from those available to students in a traditional classroom context. A central resource of a traditional classroom is the scheduled class meeting, which makes possible face-to-face interaction among instructor and students. In an internet course this is replaced by interaction via internet. In this course all time- and-place specific interaction were traded for time- and place- inspecific (asynchronous) means of communication, in order to provide students with the greatest possible freedom of when and where they did their course work. Thus, classroom interaction was replaced primarily with (1) email, allowing each student to send memos to other class members and to the instructor, (2) a "listserv", or email address where students could send memos which would be automatically forwarded to all class members, (3) an internet site, where the instructor could provide information about the course, class members, and student performance, and (4) an internet forms submission process, where students could fill out internet forms on their browsers and submit them to the instructor.

Access to and supporting information about the use of these communication devices was provided on the mg301 desktop, particularly in the second row. A prerequisite for the course was that each student have his/her own email address and the capability of sending email. It is also possible to send email and internet forms directly from a correctly configured web browser, and information about how to set up a browser to do this was provided in the third row. The second row provided links whereby students could fill out and send forms, send notes to the instructor, contribute via listserv to class discussion, and access feedback about class performance.

The second row also contained information about class members, accessed via "classmates". The only information available to nonmembers, however, was members' names. Information about individuals, including their email addresses, was kept in an area of the site known only to class members. Individual feedback was listed in a public section, but indexed by "aliases" -- here, four character nonsense words. Each individual was informed privately of his/her alias, but did not know others' aliases.

Other means of communication, such as phone calls, letters, and face-to-face meetings were also used, but only on an ad hoc and occasional basis. For example, phone calls were more frequent at the course's start, when some students were having difficulty getting their email working.

A second type of resource for the course is more familiar to classroom courses: a textbook. Though a few online courses are replacing a traditional (paper) textbook with an online equivalent or CD-ROM, there continue to be a number of advantages to a paper text, which will not be recapped here. The text, Daft's third edition (1994) of "Management", by The Dryden Press, was also used in the classroom sections of this course.

The provision of a text also means that distant learners must somehow obtain that text. Since the class was comprised of people near either our Boise or Twin Falls campuses, this was not a problem. Since our distance learning program has already established procedures for enrolling and supplying people away from these areas, provision of a text is not anticpated to become a problem.

A third type of resource is available to internet students, but not necessarily classroom students: the internet itself. The internet is in effect a kind of library available to class members. While it is unsystematic and uncharted, it is also vast and current. A part of the challenge of this course was to explore this resource and determine how it may be used to further the course's goals.

Learning Activities

Assignments were given for each week of the course. These assignments generally included:

- **Readings** from the text. In addition the instructor assigned weekly "reading memos", in which each student wrote a note to another student in the class, commenting on the reading assignment. This assignment was intended both to help students to do the readings on time, and also to encourage thinking about the comprehensibility and relevance of the readings. Requiring that the memo be sent to a peer (with a copy to the instructor) was intended to help students in their writing by placing the memo in a specific context.
- **Commentary** by the instructor on the readings and course. Each week the instructor wrote a few paragraphs on the current week's reading. These paragraphs were to some extent a replacement for a lecture. The reading might be related to the overall learning plan of the course, to the course

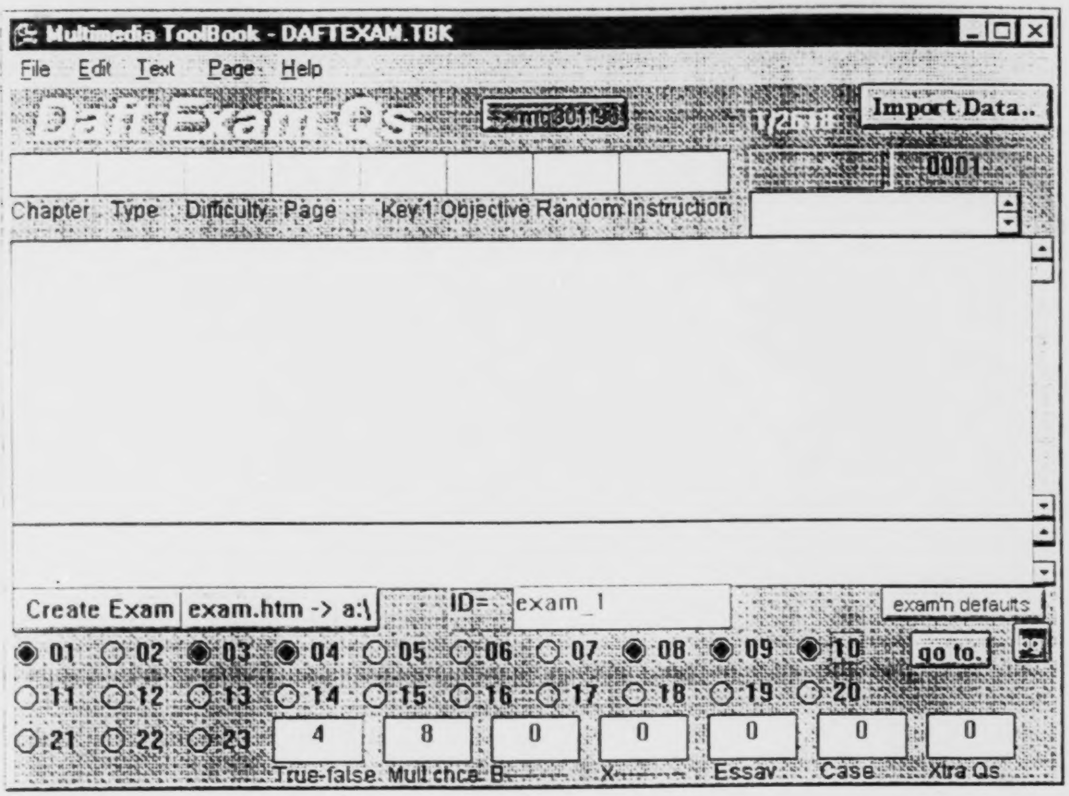
goals, and/or to larger contemporary social happenings. These commentaries might also discuss the course, how it is going, and conclude previous assignments.

- **Discussion assignments**. Every two or three weeks a discussion topic was identified, and students were asked to provide 2-3 contributions to that discussion per week. Discussions might be on an issue related to the current course material (e.g., "why have a large-scale perspective of a firm?", or "Need a manager know the business?"), on developing the course (e.g., "how can we improve our discussions?"), and on complex problem-solving (e.g., "how can firms absorb the expense of equal rights and be competitive?"). Since students often prefer to make contributions later in a discussion, 2-3 "first starters" were designated to kick off each discussion. Later on a norm of spacing contributions 2-3 days apart was established.
- **Weekly assignments**. These were short assignments intended to further the course in some way. For example:
    - reading some resource material; e.g., on setting up a browser for email or on conducting email discussions
    - filling out a "get-acquainted" form
    - filling out a global preparedness form from the international chapter, or a leadership style questionnaire from the leadership chapter
    - finding an answer to a question through internet search
    - using the internet to update a case
    - assessing organizational structures from internet organizational charts
    - determining the location of a regional "mystery" site
    - subscribing to an internet newspaper (via CRAYON)

Some longer duration activities were also included, and these are similar to those often found in classroom designs: a term paper, tests, and a final exam. These are designed to be learning activities, but with outcomes which provide indications of course goal accomplishment. They are discussed in more detail below.

Assessable Outcomes

All assignments were tracked, and some course points were allotted to their satisfactory completion. Three activities however, were specifically designed so that the student products would serve as indications of course goal accomplishment. The first of these was the traditional multiple choice exam. The exam bank of the Daft text was imported to a Toolbook book, a page of which is shown in figure 3.



0001

Chapter | Type | Difficulty | Page | Key 1 | Objective | Random | Instruction

ID= exam_1

● 01 ○ 02 ● 03 ● 04 ○ 05 ○ 06 ○ 07 ● 08 ● 09 ● 10
○ 11 ○ 12 ○ 13 ○ 14 ○ 15 ○ 16 ○ 17 ○ 18 ○ 19 ○ 20
○ 21 ○ 22 ○ 23

True-false	Mult chce	B-------	X-------	Essay	Case	Xtra Qs
4	8	0	0	0	0	0

Figure 3: A sample page from the Toolbook exam book

Programs were written by the author which would allow the selection of chapters, question types, and numbers of questions, and the generation of an exam document in the form of an html form. Since questions were drawn randomly from a bank of some 2500+ questions, it was possible to generate a large number of exams, each covering the same material, and with a low overlap of questions between exams. Thus, while it is possible that some students might save exams and pass them along with correct answers to students in the next course offering, the old exam would help very little with the next exam.

At two points in the semester, 60 question multiple choice exam forms were made available in "test 1" and "test 2" on the first row of the mg301 desktop. Students had a week to complete the exam and send the completed form back. They were advised of academic honesty rules and the consequences for breaking them.

The second assessable activity was a term paper. Since a term paper is intrinsically an asynchronous activity the assignment for this activity was taken directly from the classroom-designed course. The assignment was listed in the top row of the mg301 desktop, and links to the specific assignment. Early in the course students were asked to submit topics (just as in the classroom sections), and were given feedback as to the topic's appropriateness. Papers were submitted near the end of the semester in the body of an email message and were scored in the same way as were classroom sections.

The third assessable activity was a final exam, also listed and described in the top row of the mg301 desktop. As with the multiple choice exams, this activity was imported from the classroom section design, but was given in a "take-home" form. It was submitted via email.

The equivalency of assessable outcomes between internet and classroom sessions was deliberate, both because as a core course, there should be a strong equivalency of learning among sections, and because of interest in comparing learning accomplished between the two. Suffice it to say at this point that (1) the internet and classroom populations in this course were probably not very equivalent, (2) classroom

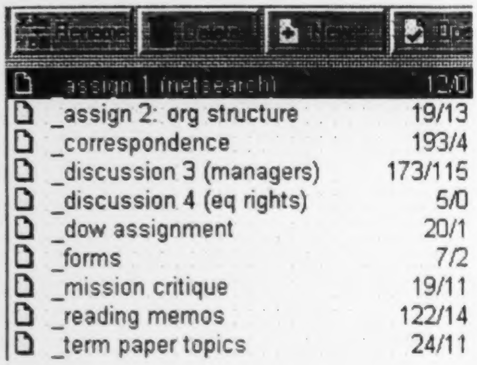
BEST COPY AVAILABLE

students did not have take-home multiple choice and final exams, and (3) the supporting resources for each group were distinctly different. These validity threats need to be dealt with before any conclusions can be drawn about equivalency of learning between internet and classroom sections.

Feedback

All student communication came to the instructor via email, whether it be memos, listserv, or forms. Given 18 students carrying out the assignments described above, it's not hard to calculate that the instructor received 100+ course-related emails a week. The first step in handling these messages was to organize them in a way which made them accessible to the instructor and to manipulation by programs.

The first phase in doing this was to organize the messages as they came in, into a series of folders. The author used Pegasus, which allowed both the creation of mail folders, and the naming of these folders. Some folders used are shown in figure 4:



```
assign 1 (netsearch)            12/0
_assign 2: org structure        19/13
_correspondence                 193/4
_discussion 3 (managers)        173/115
_discussion 4 (eq rights)       5/0
_dow assignment                 20/1
_forms                          7/2
_mission critique               19/11
_reading memos                  122/14
_term paper topics              24/11
```

Figure 4: Organization of email Folders

These folder names began with a "_" so as to place them all together at the top of the list. Each was also given an 8 character name (e.g. "discuss1"), which became the actual name of the file containing these messages on the email server. One folder was used for forms, and in general a folder was created for each assignment. Notice that there was also a "correspondence" folder, for the many memo items that had nothing to do with particular learning activities. It's very important NOT to put misaddressed or "bad" messages into these folders, since such messages usually have irregular headers which may make the analysis of messages by computer exceedingly difficult. Such messages were put in the "correspondence" folder, and a memo was sent to the originator requesting a resubmission. The general philosophy of this approach was "no deletions": all messages sent during the course of the semester were put into folders somewhere, and kept for the duration of the semester. In this way the likelihood of losing a message was reduced--as long as the mail server maintained its integrity.

In managing the class, the only folder which the instructor reviewed in Pegasus mail was the current discussion folder, since the discussion contained a sequencing not maintained in the student pages. Most of the other folders were first imported to containers on students' pages in a Toolbook book, and read there. After importation, forms data was processed by an OpenScript (Toolbook) program before reviewing. A page from the Toolbook in which this is currently done is shown in Figure 5:

BEST COPY AVAILABLE

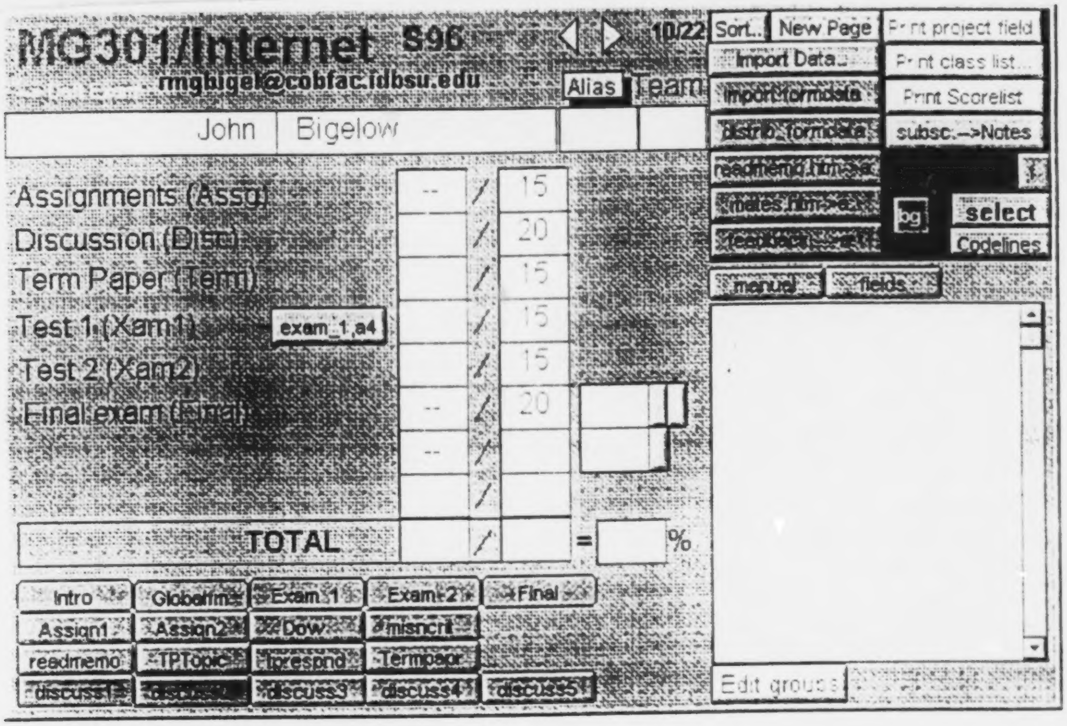

MG301/Internet S96
rmgbigel@cobfac.idbsu.edu
10/22
Alias
Team
John
Bigelow
Assignments (Assg)
Discussion (Disc)
Term Paper (Term)
Test 1 (Xam1)
exam_1,a4
Test 2 (Xam2)
Final exam (Final)
15
20
15
15
15
20
TOTAL
%
Sort...
New Page
Print project field
Import Data...
Print class list...
Print Scorelist
subsc.->Notes
select
Codelines
manual
fields
Edit groups
Intro
Globalfms
Exam 1
Exam 2
Final
Assign1
Assign2
Dow
misncrt
readmemo
TPTopic
torespnd
Termpapr
discuss1
discuss2
discuss3
discuss4
discuss5

Figure 5: Student Information Toolbook

This process worked because a person's email is accessible not only via an email program (like Pegasus), but also directly as text files. Because the folders for this course were given names, it is possible for Toolbook to to open and read them. The Toolbook page shown above is the one for the instructor. Discussion items originating from the instructor wound up on this page. A series of containers were arrayed along the bottom of the page, with names which correspond to email folders. When there is content in these containers, they are green; otherwise they are grey. When right-clicked, they enlarge so that the memos contained in them can be read. When left-clicked, a program is run which imports the corresponding email memos into the container for each student. Note that this loading was a non-destructive process for the email folder (though it cleared the corresponding Toolbook containers first), so if a mistake is made, one can simply reload the email.

The containers of the top line of this bottom row contained form data. This kind of information looks different and was treated differently. The "exam_1,a4" button in the middle of the page scored exams, put scores in the container to the right of the button, wrote an html feedback file for each individual, and created summary information about how well the class did on each question. After scoring, the instructor reviewed questions with which individuals tended to have difficulty, and decided whether the question should be counted or not. Based on this, the instructor sometimes modified the exam scoring process and reran the scoring.

After updating student pages, a program was run which created an html file which provided detailed information about what has been received from students, scores on exams (with links to detailed feedback), and a summary of how the person was doing to date. This information was publicly available from "posted scores" on the mg301 desktop, but published by "alias", as described earlier, so as to maintain confidentiality.

Feedback was also elicited from students for the purpose of improving the course. This occurred naturally, as students encountered difficulties and sought help, but in addition some short discussions on how to improve the course were also scheduled, including an end-of-semester evaluation.

BEST COPY AVAILABLE

Student Response

The course is halfway through its first semester, so more will be learned by the course's conclusion. At this point it is possible to identify both difficulties encountered by students as well as benefits which the course is providing.

Difficulties

The course was described as "pioneering", and in fact a number of unanticipated obstacles were encountered in this first round:

1. **Starting up**. A classroom "starts" according to a class schedule, and students show up at the first session. There is no equivalent event for an internet course. As a result, many students simply waited for something to happen. It was necessary for the instructor to search out/call/email a number of students, to instruct them to access the internet mg301 desktop, and to follow the instructions there.
2. **Getting up to speed**. A startup period of about two weeks was anticipated, in which students got their email and browsers working. This turned out to be about the right amount of time. While there were a few computer-ready students (these were often students with computer majors), a large proportion of the class was just beginning to establish email and browser capabilities. There was one Mac user, and one student working from campus computer labs; the rest were connecting from a home computer, either through the university's dial-up facilities or a commercial service. A wide variety of problems were encountered, some of which were new to the instructor: the listserv had been initially established as "moderated", meaning that student messages went to the instructor instead of to the listserv membership. Students seeking help from the BSU help line were told that their web browsers were "not supported". The membership roster of the listserv unaccountably disappeared on two occasions. Some people who signed up for the course apparently had not realized that it was an internet section, would not respond to inquiring calls, and subsequently dropped. People signing up from the Twin Falls campus were not initially recognized as being in the class. Early internet forms would not deliver more than the first 512 bytes of the submitted forms. Students working from home sometimes found that their phone connections had dropped while they were filling out forms. Browsers need to be configured before sending data, and some students had the frustrating experience of filling out forms, getting an error message when trying to send, and losing the data. Early feedback efforts by the instructor did not include all the materials sent by students, leading to considerable anxiety by some when they thought their assignments had not been received on time. Most of these difficulties are preventable, once they have surfaced. They are indicative of a new technology which is not yet working smoothly as an integrated system.
3. **Adjusting to the course format**. Once established, the difference between this section and traditional classroom sections became clear. Some students expressed discomfort that they could not see or directly interact with others. On the other hand, the full participation created by the process was appreciated. While the course was asynchronous there were still deadlines for completion of course activities. Some students timed their contributions to be just in time for these deadlines, in effect removing some of the asynchronicity built in to the course.
4. **Day-to-day frustrations**. The mail server went down periodically, and some students, working on a tight deadline, became frustrated when they found the mail server down and their deadline immanent. Similarly, the student working in the lab lost work when the lab went down. Our dialup facilities became nearly impossible to access between late afternoon and early morning. Thus many had to readjust their "anytime, anyplace" schedule to fit when the dialup facilities were available. As a result, many students joined commercial netservers. Access to the BSU www server has varied. At times it has been very slow, at others it has become entirely unresponsive. Listservs, which are known to everyone on the internet, have attracted "spamming". Thirty messages appeared one morning, of fictitious individuals attempting to join the class listserv. Two students received prank messages purporting to be from the White House.
5. **Submission feedback**. This is implicit in the above, but a clear issue in itself: Students attempting to send email currently do not receive feedback as to whether their messages got through. Some students would send a message several times, thinking that this would increase the likelihood of

success. This created problems for the instructor, who had to weed out the redundant messages. Particularly for exams, students would not only send the exam, but also a message to the instructor asking for confirmation of receipt. Thus, this lack of feedback created both anxiety for students and considerable overhead for the instructor. Again, this is a solvable problem, but requires someone with CGI (Common Gateway Interface) skills to produce the solution.

These difficulties have abated somewhat as we all gained expertise in conducting the course, but continue to be a significant source of frustration for many students.

Advantages

Lest it be thought that the internet course features nothing but disadvantages, students have also identified a number of positive values to the course:

1. **Flexibility**. It was anticipated that many of our non-traditional students would appreciate the asynchronous design of the course, since it would allow them to fit their course work into an otherwise filled work and family life. Several students did in fact express strong appreciation for this aspect of the course, and asserted that they would not otherwise be able to fit the course in.
2. **Creating a schedule**. It was also anticipated that the course would be attractive to some full-time students, since it would help resolve scheduling conflicts. About half the class was full-time students, and a number of them stated that they chose this section primarily because it helped them to create a semester schedule.
3. **Learning internet**. The fact that the course required use of internet was attractive to several students, who liked the idea that the course would not only deal with management, but also enable them to develop their internet skills. These students were also excited about the prospect of learning about the relevance of the internet for business.
4. **Full contribution**. It's rare to see a classroom section in which every student contributes equally; yet such contribution is instrinsic to the design of this course. Some students expressed appreciation for this, either because they were classroom contributors who felt that other class members should contribute more, or because they were somewhat shy in the classroom, and found the "public speaking" barrier to contribution removed.

In sum, the course provided both positive and negative values for students. Although many of the startup problems discussed above can be substantially ameliorated in the course's next offering, the day-to-day problems listed above don't seem to be about to disappear. Nonetheless, with some development, students in next semester's course may find its benefits to distinctly outweigh its disadvantages.

Footnotes

1. If you are reading this in paper format, the original paper is an internet html document. Underlined sections contain links which can only be accessed in the html formatted document, at http://www.idbsu.edu/business/mg/mg301/mg301pap.htm.

Bibliography

1. Christopher Newport University Online
2. The Digital Academy
3. Electronic Journals
4. The Global Electronic Classroom
5. Globewide Network Academy
6. New Ways to Learn (Andy Reinhardt). *Byte*, 51-71 (March). See also Building the Virtual College and Commentary.
7. The Online College Classroom
8. The Open University
9. Space, Collaboration, and the Credible City: Academic Work in the Virtual University
10. Teaching with the internet

11. The Virtual Classroom (UConn)
12. Virtual Courses on the Web
13. Virtual High
14. Virtual Online University
15. The World Lecture Hall

The World-Wide Market: Living with the Realities of Censorship on the Internet

Michael BJORN
Institute of Socio-Economic Planning, Doctoral Program
University of Tsukuba, JAPAN
E-mail: michael@wiz.sk.tsukuba.ac.jp

Yin Yue CHEN
Institute of Socio-Economic Planning, Doctoral Program
University of Tsukuba, JAPAN
E-mail: chen@wiz.sk.tsukuba.ac.jp

Abstract: With the passing of the Telecom Bill containing the Communications Decency Act (CDA) on February 8th, 1996 in the USA, the openness of communication on the Internet was put into question. Even though the CDA was later deemed unconstitutional by a federal court, the question remains open, and other countries may take similar action.

For this reason we have implemented a system built on a classified ads structure which is safe for the provider who is faced with the problems of responsibility for on-line material, decisions regarding what kind of material is unsuitable, and multiple language monitoring -- while at the same time providing a sufficient degree of communicative freedom and perceived value-of-service to the user.

1. Introduction

When U.S. President Clinton signed the Telecom Bill containing the Communications Decency Act (CDA) on February 8th, 1996, many home pages on the WWW turned black and were accompanied by a link, asking: Why is this page black? The answer is simple: Fear that banning "indecent" material, would also hamper the unpreceded possibilities for masscomunication on the Internet. Traditional mass-communication takes the form of a one to many relationship; one radio or TV-station broadcasting to millions of people, one newspaper or magazine with millions of subscribers, etc. The Internet, on the other hand, allows for all imaginable relationships between senders and receivers of messages: one to one, one to many, many to one, many to many.

This multitude of communication possibilities makes it very difficult to make laws that are effective or have the intended effect. Another complicating fact is that the Internet is a truly global system, and since there is no globally binding legal system, lawmakers on the national level might feel pressed to take more drastic action. Companies naturally fear that they will be struck by new laws in ways that they could not foresee and may for this reason hesitate to create new Internet-services. In this paper, we take the stand of a prospective service provider, and ask ourselves if it is possible to build services for the World-Wide Web that will not cause the provider of the service legal problems, while at the same time having potential to generate income and be of value to users.

We believe that we have found one possible framework that satisfies our requirements. We here present a system which:

- Preserves all communication relationships mentioned above. This implies that the owners of the service and the creators of information on the service are not necessarily the same.
- Lets the manager of the service clearly decide how responsibility of what is published should be divided between the service provider and the users of the service.
- Gives the manager full control of what is published on the service.
- Does not have to be moderated for "indecency" or otherwise unsuitable material.

2. Problem Areas

We have identified three problem areas which we have to deal with if we want to implement a system which minimizes legal risks for the provider of an on-line service. These problem areas are described below.

2.1 Responsibility for On-Line Material

In traditional communication channels, the question of legal responsibility as well as copyright was easy. The publisher had the legal rights and that was it. But with computer networks it is more difficult, and we have to make a distinction between content provider and content creator. CompuServe, for example, is a content provider in the sense that it distributes e-mail on its network. But is it responsible for all the content in those e-mails? Traditionally, the answer would be yes, because the company has designed the service and also has subscribers who pay for its services, just like a newspaper or magazine. However, network users contribute their own material and thus share the responsibilities for (and rights to) the content with CompuServe.

As an example of this, CompuServe in the winter of 1995 was accused by criminal prosecutors in Germany of providing access to sites which contained indecent material [Wall Street 96]. Even though CompuServe was not the creator of the content in those sites, it acted upon the accusations, in a sense accepting responsibility for the content.

2.2 Definition of Characteristics of Unwanted Material

Another problem is how to define what kind of material laws are supposed to protect against. The Communications Decency Act uses the rather vague term "indecent" material. For this reason, the law on June 12 was deemed unconstitutional by the federal court of Philadelphia [Quittner 96]. This injunction prevents any prosecution pending appeal to the U.S. Supreme Court -- but the problem is likely to resurface again in other countries, no matter the outcome this time.

However, as long as we have difficulties to properly define what we want to block out, we have difficulties providing tools that will do the job. Currently there are two methods, the first takes the approach that cutting off too much is better than not cutting at all; a list of "suspect" sites is maintained and those sites are blocked out completely. The second method uses filtering software to scan and censor texts with "unwanted" words.

The first method is unsatisfactory because it may block out useful and/or harmless information as well, and the second method faces a language problem: Human language is complex and words which are harmless in one context may indeed be indecent in another.

As an example where both methods were used, CompuServe totally blocked out some 200 sites following allegations by the German criminal prosecutors that these sites contained unsuitable material. Following customer complaints about the total block-out, the company then licensed CyberPatrol, a software package that automatically restricts access to indecent material by blocking sites and filtering content of individual files [Investor's 96].

2.3 Multiple Language Control and Language Ambiguity Handling

The USA as well as many European and Asian countries today have populations with a mixed composition of cultures and languages. For this reason, interaction on the Internet will take place in many languages, nationally as well as internationally. This makes it quite difficult when trying to block out unwanted material. An example could be when a mailing list is used by a group of people speaking a language unknown to the list moderator. The list moderator, then, has no way of knowing whether the discussions in that language are indeed within the scope of the mailing list, if they are just a lot of empty talk, or if they are in some way unsuitable, potentially endangering the company responsible for the mailing list.

To continue our CompuServe example, the Cyber Patrol software was initially offered in English and German versions, with French and Spanish versions to follow. How effective this multi-language filtering-approach will be is difficult to say, since there are at least three additional complexities when trying to adapt filtering software to multiple language environments:

- First of all, there are cultural differences as to what is to be considered indecent. As an example, we can see how a similar problem is handled in a different medium, the movies. American films are often quite heavily censored when shown in Europe, because of a high level of violent content which is considered perfectly acceptable in the USA. On the other hand, European films are often re-cut or X-rated in the USA because they contain too sexually explicit scenes for the general public. What is indecent then, in the sense that it is considered not suitable for minors or the general public, varies between cultures. Similarly, filtering software must be customized for different types of scanning in different languages.
- Secondly, harmless words used in a non-English language may be blocked out by English-based filtering software such as SurfWatch, CyberPatrol and NetNanny. As an example, consider the word 'sex', which is part of the standard filter-out list in CyberPatrol -- in Swedish it can simply mean the number '6'. Other words in the CyberPatrol list which are incompatible with Swedish are 'love' which is the name of a famous Swedish author, and 'sade' which is the past tense of the verb 'say'. Other languages have similar incompatibilities.
- Finally, the filtering situation gets even more complicated when the subject who is to be blocked out speaks two or more languages. It is a fact that many young people across the world today learn and speak more than one language. Filtering software must then be able to handle combinations of filters for different languages.

3. System Solution

Our system, called the World-Wide Market, handles transparent transfer of information between as well as within different language groups so that interactivity is preserved, while at the same time avoiding the above problems with unsuitable material. The three problem areas described above are addressed in this section.

3.1 Scoping of Service -- to Clarify Responsibility for On-Line Material

The system uses scoping to give the service provider full responsibility and control over what is published on the system. Thus, the service provider can decide what level of "indecency" is acceptable, and easily change that level in accordance with changes in regulations, etc. This amounts to a voluntary restriction, which may not only be of interest to a provider because of fear for legal action. Other motivations include that the provider wants to focus the service and reduce the amount of processor time used for unrelated activity, to qualify for the PICS (Platform for Internet Content Selection) rating system, etc.

The system is based on an extended version of the classified ads sections found in many newspapers. The scope of such a system consists of actions, intentions and sections as in the following example:

- Actions : (Register, Insert, Delete, Update, Search, Browse)
- Intentions : (Sell, Buy, Donate, Receive Donation)
- Ad Sections : (Automobiles, Books, Computers, Electric Appliances, Music Recordings, Real Estate, Rental Contracts, Video Games)

Users of the system communicate by establishing relations between several enumerations, in effect building a subset with items from the involved enumerations. They are free to communicate anything within this scope, but they have no way of going beyond it. A user could for instance select the subset (Search, Sell, Automobile), in effect formulating the sentence: "I would like to search the sell ad listings for an automobile."

The scope can easily be extended by for instance adding other sections to include dating service, pen-pals, offer of services instead of products, etc.

3.2 Enumeration of Vocabularies -- to Define Characteristics of Acceptable Material

Instead of concentrating on what it is not acceptable to say, we take the much easier approach of deciding what it is perfectly alright to say. We do this by implementing sets of pre-specified topics and vocabularies which are strictly enumerated and necessary to make meaningful statements in the above sections. We thus avoid the pitfalls and complexities of huge free-form text translation systems. Problems with judging performance of such systems [King 96] also make us wary of that approach. The number of possible combinations of enumerated vocabularies in our system is quite large, and provides sufficient, language independent information to potential buyers and sellers of products that can be traded within the system.

We currently have no scientific criteria to show that our selection of topics and vocabularies is indeed necessary and sufficient to provide meaningful communication in specific section. Our approach here has been to make empirical investigations of classified ads in newspapers as well as on-line services, and include the most commonly found topics and their related vocabularies. We are considering using computerized information extraction from on-line services, such as in the FRUMP system, [DeJong 79] and [DeJong 82].

To continue the above example, the section Automobile may have the following definitions:

- Topics : (Vehicle Type, Manufacturer, Price)
- Vocabularies : ((Sedan, Hard Top/Coupe, Open Convertible, Sports Car ... etc.) (Acura, Alfa, Audi, Austin, BMW ... etc.) (1 .. 20,000))

The user can now be more specific in his formulation: "I would like to search the sell ad listings for an automobile which is a sedan or sports car made by Audi, costing no more than $ 10,000."

The mathematical foundation for the approach in 3.1 and 3.2 above is a simple application of combination theory. The system is a combination of well-formed sentences which can be gotten out of the formula:

$$\textit{Actions} * \textit{Intentions} * \textit{Sections}$$

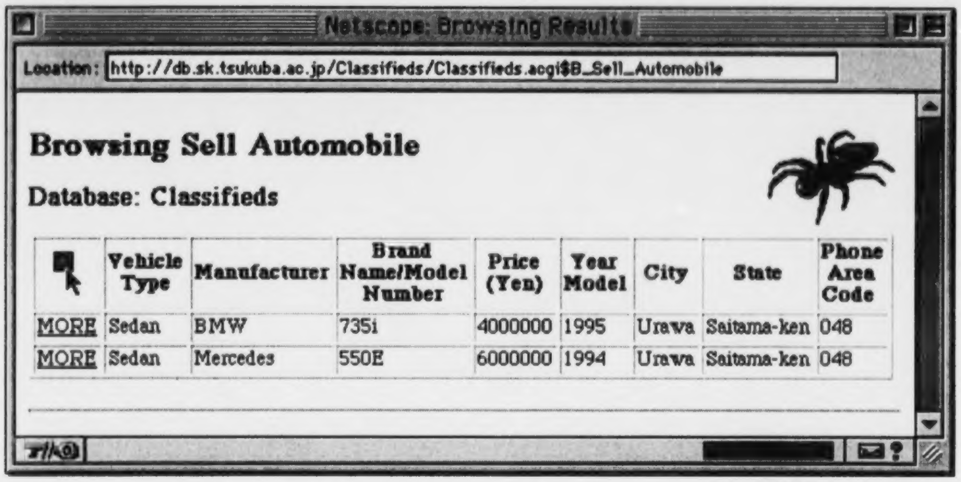
where each Section has the following number of possible combinations:

$$\textit{Vocabulary-Topic1} * \textit{Vocabulary-Topic2} * \ldots * \textit{Vocabulary-TopicN}$$

During the design of the system we have come to talk about our enumerated language as a "pickup-language". This is because we provide a graphical, mainly menu-based interface for submitting and trading of goods in each section, as opposed to newspapers which require more human processing, especially for ad submissions. For example, if someone wants to sell a car, he/she inputs information by picking up menu choices for maker, brand, model etc. Display of information to potential buyers is then based on menu selections, so that unsuitable information neither can be in- or output from the system.

3.3 Mapping Between Equal Vocabularies -- to Handle Multiple Languages

So far, we have mostly talked about input of information into the system, using our "pickup-language". However, for communication to occur, there must also be output. All input to the system is stored as English "pickup-words". When someone browses the system, this information is then presented in tabular form, as shown in Figure 1. This tabular presentation is easily readable and has the advantage that it is highly language neutral and does not depend on other syntactic rules, such as word order, plural forms, etc. A mapping for each word is provided so that the user can browse the system in his/her language, regardless of what the input language originally was [Bjorn and Kim 96].


Netscape: Browsing Results
Location: http://db.sk.tsukuba.ac.jp/Classifieds/Classifieds.acgi$B_Sell_Automobile

Browsing Sell Automobile

Database: Classifieds

	Vehicle Type	Manufacturer	Brand Name/Model Number	Price (Yen)	Year Model	City	State	Phone Area Code
MORE	Sedan	BMW	735i	4000000	1995	Urawa	Saitama-ken	048
MORE	Sedan	Mercedes	550E	6000000	1994	Urawa	Saitama-ken	048

Figure 1: Tabular representation of the "pickup-language"

The same is also true for input: For each menu choice, a mapping from a base language to each supported language is provided, so that the user can interact with the system in his/her preferred language but still make the information available to speakers of other languages. A mapping (from the topic level down) for the above example list shown in Table 1. By performing direct one-to-one mapping between languages, we avoid all problems regarding filters with different impact on different languages, as well as context dependencies and other semantic problems.

USA	Japan	Taiwan
Vechicle Type	自動車の種類	汽車之種類
Sedan	セダン	小轎車
Hard Top/Coupe	ハード・トップ／クーペ	轎跑車/三門轎跑車
Open Convertible	コンバーチブル	敞篷跑車
Sports Car	スポーツ・カー	跑車
Station Wagon	ステーション・ワゴン	旅行車
Mini Van	ミニ・バン	迷你箱型車
Van	バン	箱型車
Recreational Vehicle	RV・カー	野外休閒車
Camping Car	キャンピング・カー	露營車

Table 1: Mappings for English, Japanese and Trad. Chinese

4. User Value

Restricting communication in the way we have been doing might be considered practical from the service providers point of view, but why should this rather Orwellian approach appeal to end-users, who are free to choose systems with more expressional power? Apart from the language-independence which obviously benefits the user, we see a number of other reasons.

- **Ease of use**: Since much of the information to be input to the system is already present in menus, the user does not need to type very much, worry about personal style or formulations, etc.
- **Less decision making**: Since the structure is standardized, the user does not need to decide what information needs to be input. He/she can also rely on passive knowledge when it comes to maker

name, brand name, etc.

- **Personal privacy**: With the advent of more and more sophisticated query mechanisms on the Internet, many users are concerned about how bits and pieces of information that they supply can be used to make inferences about them. Our system can only provide standardized information and should be less of a threat to personal privacy.
- **Harassment**: Considering that it is relatively easy to forge e-mail addresses and IP numbers, a user could easily be connected to a sexually explicit or otherwise harmful statement in a free-form text system, which can lead to harassment or other inconvenience for the user. In a system with low expressional power, this is less of an issue.
- **Matching service**: Since the information input to our system is highly structured, it is very well suited for computer analysis. For this reason we can provide services that free-form text systems can not provide, such as parameterized matching of buy and sell ads with notification of the match to the advertisers.

5. System Implementation

The system is strongly typed for information about goods in pre-specified categories. It consists of a relational DBMS, a WWW-server, as well as a Common Gateway Interface (CGI) application together with a set of HTML files for each language, building on a system described in [Bjorn and Hotaka 95]. The CGI processes language mappings and other meta-data, handles conversion between HTML data sent by users and SQL used for database queries, as well as builds dynamic HTML pages in reply to the users' actions [Bjorn 95].

- **Meta Data Structure**: Language mappings are stored in a meta-database, so that the system can automatically map information for display into the language preferred by the user. These language mappings also contain MIME headers for automatically changing the browser's display format into the preferred language.
- **Extensibility**: The system is easily extensible: When a new language is added, only the set of HTML files as well as a resource file for the CGI has to be translated -- no changes have to be made to the database schema; instead, the meta database is simply updated with the new language mappings. The computers used as servers do not even have to support the language in question -- the only requirement is that the user's browser and computer support the desired language.
- **HTML limitations**: HTML was originally designed as a markup and linking language, but has been extended with various features which it sometimes lacks the expressional power to support satisfactorily.

  Our major problem was the FORM construct, which is not primarily used for linking or markup, but for data input and query in a way more reminiscent of database languages than of "page-layout" languages. The FORM construct only accepts text input, has HTML has no typing constructs. We rolled our own very simple typing system by reserving field names for certain data types, such as integers and sets. This is only mentioned here to point out a fundamental shortcoming with HTML.
- **Extent of implementation**: At present, the system contains English, Japanese and Traditional Chinese. These languages are fully transparent to each other, apart from vocabularies relating to person and place names. We plan to extend the relational database with such vocabulary mappings in the future.

A strictly R&D only version of the system can be accessed at: http://db.sk.tsukuba.ac.jp/classifieds/classifieds.html, and a production version an be accessed at: http://wholecycle.eccosys.com/

6. Conclusion

With the passing of the Telecom Bill containing the Communications Decency Act (CDA) on February 8th, 1996 in the USA, the openness of the Internet was put into question. However, at the same time it is very difficult to control what kind of material is actually transferred on the Internet. For this reason, we have designed a classified ads based system, implementing a restricted pickup-language which can be used for trade of products or services or exchange of other types of information. Although free communication between buyers and sellers is not supported, we provide essential, multilingual product information to make the buyer/seller aware of a certain business opportunity, and establish an incentive for the buyer/seller to make an effort (e.g. via e-mail) to conclude the deal.

In the future, we would like to extend the system to include other categories of products and areas of information, as well as provide a better theoretical base for the extraction of sufficient and necessary information from human languages.

7. References

[Bjorn 95]. Bjorn, M. (1995). An Interactive Relational Database Gateway with Load Balancing. AUUG95 & Asia-Pacific WWW Conference, 1995, Australia, AUUG'95 & APWWW'95: 263-272.

[Bjorn and Hotaka 95]. Bjorn, M. & Hotaka, R. (1995). An HTTP Gateway for Interactive Relational Database Query with Deferred Reply. IEICE Transactions on Information and Systems, Vol. E78-D, No. 11: 1407-1411.

[Bjorn and Kim 96]. Bjorn, M. & Kim, H. (1996). Learn English or Else - Commercial Interaction on the WWW in Asian Languages, Information Retrieval in Oriental Languages, 1996, Taejon, Korea, KORDIC: 57-63.

[DeJong 79]. DeJong, G.F. (1979). Prediction and substantioation: A new approach to natural language processing, Cognitive Science, No. 3: 251-273

[DeJong 82]. DeJong, G.F. (1982). An overview of the FRUMP system, Strategies for Natural Language Processing. Erlbaum, Hillsdale, N.J: 148-176

[Investor's 96]. Investor's Business Daily, 14 Feb 1996, A9

[King 96]. King, M. (1996). Evaluating Natural Language Processing Systems, Communications of the ACM, Vol. 39, No. 26: 63-72

[Quittner 96]. Quittner, J. (1996). Free Speech for the Net, Time, June 24, 1996, Vol. 147, No. 26, 52-53

[Wall Street 96] .Wall Street Journal, 5 Jan 1996, B2

A Broadcasting Company goes Internet

Manfred Bogen <Manfred.Bogen@gmd.de>, GMD, Germany

Michael Lenz <Michael.Lenz@gmd.de>, GMD, Germany

Susanne Zier <Susanne.Zier@gmd.de>, GMD, Germany

Abstract: This paper describes the experiences made in bringing the first German broadcaster, Deutsche Welle, to the Internet. The approach presented is feasible for other institutions and companies starting with Internet and WWW.

An integration of different value-added services like WWW, electronic mail, FTP and WAIS was urgently needed. For cost, risks and speed reasons two external partners supported the Deutsche Welle Internet task force: one for the layout and the information design and one for the provision of the Internet connectivity and the operation and the enhancement of the information service.

A properly defined business model, a guaranteed quality of service level and a secure access to Internet was as crucial as statistics since the service start .

1. Introduction

The time where Internet was a big playground for adults and children is almost over. Individuals, organisations and governments use the Internet for a wide range of professional and commercial activities. Among them are local, national and international broadcasters, TV and radio stations competing each other. The Internet and especially the Internet application World Wide Web (WWW) is an ideal tool for them to deliver their program: a world-wide coverage, multilingual and multimedial, and finally cost-effective. The first obstacle against the use of this new medium is a quick and good start. External help is warmly welcomed. This was also the plan of the *Deutsche Welle (DW)* to enter the Internet arena.

In the first chapter after the introduction we shortly present Germany's international broadcaster Deutsche Welle and its program. The next two chapters describe the first common plan to have an Internet-based program, how it was overridden by the Internet reality only three weeks after the service start and how a more general plan has been set-up among the co-operation partners. The last chapter heading the conclusion shows our experiences made in terms of statistics and quality characteristics. The conclusion sums up our lessons learnt on our way to move the services to the Deutsche Welle responsibility.

2. Deutsche Welle

Deutsche Welle is Germany's international broadcasting service [Deutsche Welle 95]. Regular short-wave transmissions started on May 3, 1953. Through its radio and television programs, Deutsche Welle provides listeners world-wide with a comprehensive picture of political, cultural and economic life in Germany. Deutsche Welle is chiefly financed with funds allocated by the Federal Government. Its annual budget for 1994 totalled approximately 600 million Deutschmarks.

In 1994, the management of Deutsche Welle decided to complement the traditional program branches radio and television with new media. Its major objectives were to keep the multilingual program and the world-wide coverage, to improve its information capabilities, and to allow a more flexible and fast communication with the audience as soon and fast as possible. A small internal task force (subsequently called DW-Internet) was set up.

Internet applications like World Wide Web (WWW), file transfer (FTP) and electronic mail (email) seemed to qualify to achieve the goals mentioned. Other broadcasting companies like 'BBC' and 'Voice of America' had already started on the Internet as information and program providers. They have a similar mission like Deutsche Welle. However, taking a pure business approach there was a certain uncertainty

- whether these new applications could really satisfy the new goals of the broadcasting service,

- whether it could be possible to attract new audience world-wide,

- about the amount of money to invest for personnel, servers, and bandwidth,

- whether a return of investments would be guaranteed,

- and whether it should be a branch of its own or part of the radio or TV program.

Therefore the risks for the company should be minimised.

3. The Deutsche Welle General Plan

A traditional business-oriented way to solve this problem is outsourcing. Another important point is a clear distribution of responsibilities and work. Setting up Internet services require networking and system administration knowledge as well as skill in organising and preparing the data [Liu et al. 94].

Therefore two external partners were sought to co-operate with the DW-Internet task force being responsible for the program itself and the raw material:

- one co-operation partner for the layout and the information design in the World Wide Web.

- one co-operation partner for the provision of the Internet connectivity and the operation and the enhancement of the information service.

Both co-operation partners should be experienced, should be willing to co-operate and should be capable to act as a think tank for the Deutsche Welle in questions related to Internet, Internet applications and future directions. The personnel of Deutsche Welle being responsible for networking and communications should have an opportunity to learn in practice the management of the new Internet-based applications in order to be enabled to run the service needed by themselves in medium terms (migration plan).

Apart from that, the service should come up as fast as possible with a minimal amount of bureaucracy (e.g. procurements and registration of a second-level domain) but should be operated under professional conditions with guarantees for quality of service and a certain amount of security.

3.1. The Co-operation Partners

GMD is the German National Research Center for Information Technology [GMD 96]. In its *Media Communication Institute* it performed the engineering of value-added services like e-mail, file transfer, and directory services for different co-operation partners. Especially in this context WWW-based developments, integration of services, and works related to Secure Internet are of major importance. GMD qualified to be the co-operation partner for the service operation.

gekko is a small, spin-off company of GMD located in the GMD premises called *GMD Technopark*. As a combination of a traditional consulting and advertising agency with a general purpose technical service provider gekko specialises in the provision of access methods as well as the development of ways of representation and turn-key information and advertising systems for companies wishing to establish their presence on the Internet [gekko 96]. gekko qualified to be the co-operation partner for the information design and layout.

3.2. The Time Schedule and a First Plan

DW wanted to be present on the Web. A small, low-budget project based on existing hardware - the so called introduction phase for setting up the basic contents, introducing the services and merchandising - was defined starting in September 1994 with expected end in August 1995. Afterwards, depending on the results of this phase it should be decided whether and how to continue.

In order to give DW's Internet start a sound basis careful planning was crucial. Within a couple of meetings we defined the realisation steps as well as the responsibilities and the distribution of work. We also identified some basic guidelines. The realisation must be scaleable, i.e. the system must be adaptable as the popularity and hence the number of accesses grows. The solution must be extensible so that new services could be integrated as they become popular. And last but not least the migration of the system from GMD to DW was a major aspect from the very beginning.

4. The Internet Reality

In this early phase we only realised WWW as the most essential service and beside a regular data backup almost no quality of service guarantees were given. By October 1, 1994 DW was on the Web! The service was announced on different national and international mailing lists and Web indexes. In addition DW started advertisement in its radio and TV programs.

For a fast and less-expensive solution for an early start of the service we used the second-level domain name 'gmd.de' for DW's Internet service names. The machine was directly connected to our local network and DW staff was able to connect via modem to the server machine.

But already three weeks after going on-line all planning was passed by reality. DW's on-line information at the evening of the ´94 election of the German parliament (Bundestagswahl 1994) was an overwhelming success: 35,000 accesses at that day and another 30,000 during the next week exceeded the most audacious expectations by far. The service machine already got to its limits. So following our contract we had to talk to each other again.

4.1. Services Needed

People at DW were very impressed and motivated by the results of the early phase, especially by the events around the Bundestagswahl. DW decided to finish the introduction phase immediately and to continue with a regular production phase. It was our task now to map the additional requirements to feasibility with respect to hard- and software as well as bandwidth limitations and costs. At first they had wanted to be present in Internet at all, now they wanted more. Finally we agreed upon the essential services

- WWW,

- FTP,

- WAIS and

- email.

Beside on-line browsing within the Web, FTP was needed to give the users the ability to download large files like e.g. DW's monthly program schedule. For large Web sites like DW an index is necessary to be able to search for information. We decided to use WAIS as the standard indexing service within the Internet. But DW didn't solely want to provide information, they were also interested in direct user feedback. Thus email was also needed. Also these services had to be integrated in an efficient and useful manner.

4.2. Quality of Service

For the production phase effective quality of service guarantees were clearly needed. We agreed upon the guarantees and upper limits for the

- operation of the services (24 hours a day, 7 days a week),
- reaction upon any failures,
- backup of the data,
- system resources, and
- connection bandwidth.

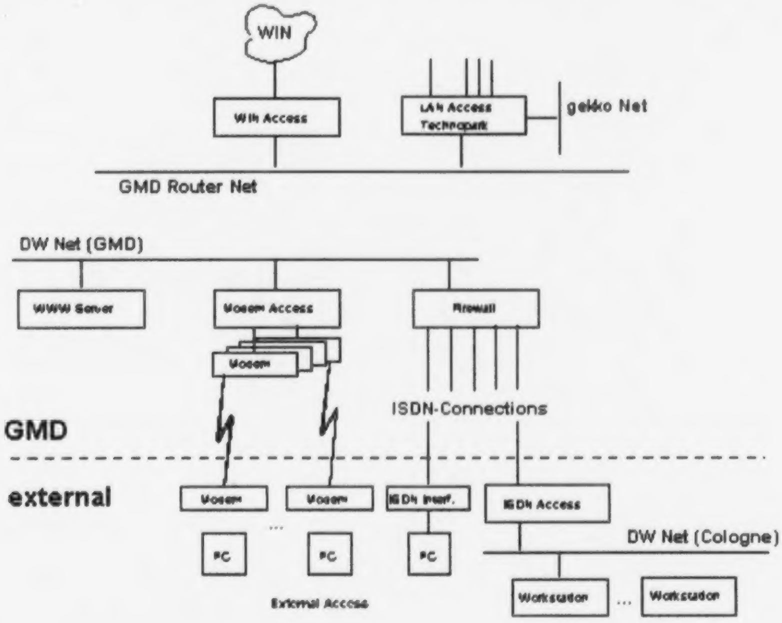
In addition we agreed upon monthly usage statistics for DW and quarterly management reports of the current status and the ongoing developments.

An effective mechanism for announcing any kind of failure and the persecution of its processing and eventual elimination became necessary. Our solution relies on email and mailing lists. Different roles like e.g. *ftpmaster* or *webmaster* are defined and related to respective mailing lists. Internally we clearly defined the responsibilities of the involved system administrators (machine, network, services) for reactions upon announcements.

In addition, in order to guarantee the quality and to talk about all existing or arising problems we (GMD, DW and gekko) meet on a regular basis. At the beginning of the co-operation we meet more frequently to fix all the details and to give people at Deutsche Welle the necessary Internet background. Later on the intervals between the meetings became longer.

4.3. The Production Phase

The production phase started January 1, 1995. According to the results of the introduction phase it was immediately clear that we need a much more powerful machine for this phase. But to minimise costs as well as administration overhead we continued to run all services on a single machine which was sponsored by GMD. We placed the machine on a separate service net.


WIN
WIN Access
LAN Access Technopark
gekko Net
GMD Router Net
DW Net (GMD)
WWW Server
Modem Access
Firewall
Modem
ISDN-Connections
GMD
external
Modem
Modem
ISDN Interf.
ISDN Access
...
PC
PC
PC
DW Net (Cologne)
External Access
Workstation
...
Workstation

BEST COPY AVAILABLE

Figure 1: System structure

Via a firewall system this service net is linked to GMD's IP backbone [Chapman and Zwicky 95]. As a member of GMD's Technopark gekko is also connected to the central backbone. Beside the service machine we installed a modem pool at the service net which users can connect to for retrieving information from the servers. Our infrastructure is completed by an ISDN router which is connected to GMD's backbone. One of its interfaces is reserved for use by DW only to transfer raw text material from DW to gekko for conversion into HTML as well as to give some selected employees of DW full Internet access via GMD.

We used standard software (NCSA's *httpd*, *ftpd* from the Washington University in St. Louis, *freeWAIS* and *sendmail* [NCSA 95, Liu et al. 94, Costales et al. 93]) for realising all services due to their rich functionality and widely usage within the Internet community. Especially, we used a public domain Web server because no matured commercial server was available at that time.

Also the migration to our new server machine at the beginning of the production phase was smoothly and nearly transparent to the users. One service after the other was transferred by changing its DNS entry after preparing the service on the new machine. At the beginning, some users and programs (especially Web crawler for indexing entire Web sites) continued to use the old IP address. Thus, for a transitional period we kept the old servers running but pointing any users to the new machine.

To guarantee and to verify our quality of service agreements we monitor the hardware with our internal SNMP-based network management system [Eisenblätter et al. 93, Case et al. 1990]. Using these mechanisms we are also able to detect any violations of agreed resource limits (like e.g. Internet bandwidth) and to obviate them if necessary. For the beginning we also manage DW's ISDN connection to GMD. We keep access data for all services as well as the router utilisation in order to produce our monthly usage statistics. Backups of the data areas are made automatically over night.

4.4. Security

According to its mission Deutsche Welle contributes to the view of Germany in foreign countries. The content of the services therefore has a somewhat official image so that its integrity has to be ensured. On the other hand the Internet isn't secure today. Thus, considerable security efforts has to be undertaken.

Placing DW's server machine on a separated network protected by a firewall [Fig. 1] allows us to use IP filtering and access lists to enhance security [Chapman 92]. To improve the security of our services even further we used some host and network security scanners [SATAN 95]. We also permanently monitor the activities of the machine in order to detect any unusual behaviour which might be a clue to hacker activities. In addition for the ISDN connection between DW and GMD additional filtering for protocols, ISDN numbers of the sending and receiving sites etc. are activated. GMD's internal networks are also secured by additional security mechanisms. Thus DW's net is secured against attacks from the Internet and the server machine as vice versa GMD against attacks from the server machine and over the ISDN connection.

4.5 How to Make Money Out of it

DW's first intention was not to make but to save money by its Internet activities. Instead of sending program schedules via ordinary mail to recipients all over the world users can access these information on-line by WWW or download them via FTP. Beside the cost savings for Deutsche Welle customers get a much faster and more actual access to the data.

By analysing the server's log files, the usage of the information and the accesses per country and language can be determined quit easily. One must know that it is not possible for international radio and TV broadcasters to get real numbers about their listeners and viewers and by that about the acceptance of their program. But this information is absolutely necessary to adapt the content properly to the users' needs.

In addition, Deutsche Welle profits from a synergy effect, e.g. by the co-operation and the exchange of material between its different branches and editorial departments and by advertising for each other. Last but not least, Deutsche Welle is funded by the Federal Ministry of the Interior and due to this financial backup a short term return of investment has not to be their most important goal. But nevertheless Deutsche Welle is currently thinking about requesting money for special services like e.g. access to their archives.

5. Experiences

In our co-operation with DW we made several experiences about quality characteristics of Web servers and training novice users of the Internet.

5.1. Statistics

Statistics are important for the Deutsche Welle as well as for GMD. Our contractor has to report on the results of the co-operation and has to justify the investment in Internet internally. GMD uses these numbers for quality control, monitoring and capacity planning.

Figure 2 shows the usage of the different services since the beginning. The most important service for Deutsche Welle is the Web server. During the first months the number of requests was rather stable - around 350000. Since October 95 parts of the program of the *Süddeutsche Zeitung*, one of the most popular German newspapers, are located on the Deutsche Welle server too. As a result the number of requests grows up to 600%.

FTP and email services were installed at the begin of 1995. Compared with the Web server the utilisation of FTP and email is small. The total traffic in June '96 was 6.2 GB, 6 GB for Web traffic and only 140 MB for FTP and 9 MB for email. Despite the low traffic volume email is important for Deutsche Welle to get feedback from its Web readers and its radio and television audience. It is also important for GMD and gekko to get end user error messages.

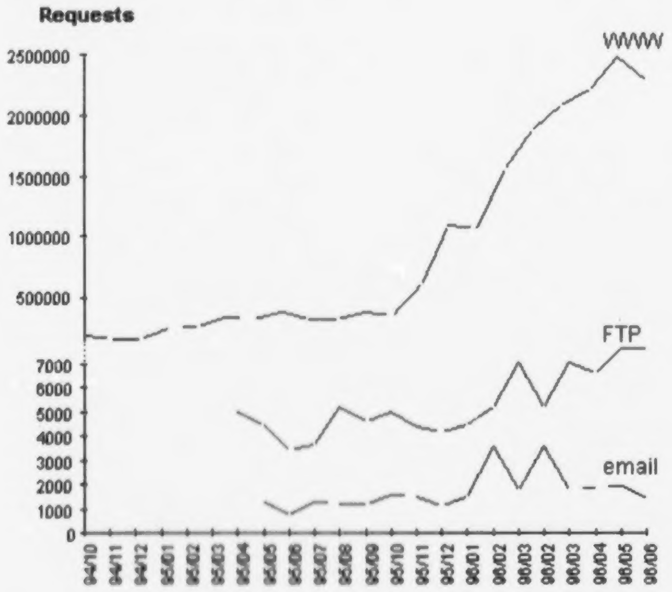


Figure 2: service usage

5.2. Quality Characteristics and Consulting

BEST COPY AVAILABLE

With respect to our quality characteristics [Quality of Service] the average CPU utilisation is around 30%. Even for a Web server with such a number of requests this is a high value. It is caused by the dynamic creation of HTML documents. Most of HTML pages are individually constructed from a *msql* database. *msql* is a public domain database software. For the costs of a higher CPU utilisation the advantages of using a database is the high flexibility in serving different user interests and an easier way to manage the Web content [Baker et al. 94]. The disk utilisation is smaller than expected because DWës Web offering only contains a limited number of multimedia files. But this will change in the future. The used bandwidth increases steadily.

In addition to setting up the server machine and to maintaining the services we had to support the external users as well as the DW employees. Normally the Web users send error messages via email to accounts like *webmaster* or *ftpmaster*. Many problems depend on user and configuration errors on the client side. If possible we offer help for solving these problems. Questions about the contents and representation of the pages or problems with *cgi*-scripts are solved by gekko.

To prepare the server migration, we have to train Deutsche Welleës system managers. Beside knowledge about the software, its configuration and the requested system resources a basic understanding of the Internet is crucial. In addition to training the administrators we have to support the information providers and users within DW - especially the editors. Traditionally they have no experiences with the Internet and its applications (like FTP and email). They need advisory how to configure the client software and to use the services. Further more they need support about security in general, personal email accounts, and secure passwords.

6. Conclusion: The Future

The choice of two external co-operation partners speeded up the process and minimised the risks and costs for the Deutsche Welle due to the Internet and WWW experience of the two companies involved which by themselves also learnt a lot [The Internet Reality]. The distribution of responsibilities has proved useful in the day-to-day business.

In spite of the pressure from the market and the competition it is very important not to start in a hurry but to do a careful planning of the activities and the service to be offered. Once you have started a Web-based service you have to continue it in a professional manner which is much more work-intensive and difficult than just coming up with a Web server.

It became obvious that WWW is not enough for an international broadcaster and that an integration of different value-added services like email, FTP and WAIS was urgently needed to fulfil the needs of the Deutsche Welle and its audience. A properly defined business model, a guaranteed quality of service level and a secure access to Internet was as crucial since the service start as statistics. The approach chosen proved feasible for other institutions and companies starting, sometimes as co-operation partner of GMD, with Internet and WWW.

7. References

[Baker et al. 94] Baker, T., Brüning, I., Klein, L., Michael Lenz, M. (1994). Starting a Web site: issues of quality control. Computer Networks and ISDN Systems 26 (Suppl. 2 double issue), 55-61, Elsevier Science B.V., Holland.

[Bogen et al. 94] Bogen, M., Hausen, H.-L., Worst, R. (1994). Handling of QoS Characteristics. Computer Networks and ISDN Systems 26 (Suppl. 2 double issue) , 107-118, Elsevier Science B.V., Holland.

[Bogen et al. 95] Bogen, M., Hausen, H.-L., Worst, R. (1995). Quality of Service in a Multimedia Environment. NETWORLD + INTEROP 95, Engineer Conference Notes, Las Vegas.

[Chapman 92] Chapman, D.B. (1992). Network (In)Security Through IP Filtering. Proceedings of the Third USENIX UNIX Security Symposium, Baltimore.

[Chapman and Zwicky 95] Chapman, D.B., Zwicky, E.D. (1995). Building Internet Firewalls. Sebastopol, CA: O'Reilly & Associates Inc.

[Cheswick and Bellovin 94] Cheswick, W.R. , Bellovin, S.M. (1994). Firewalls and Internet Security. Addison-Wesley Professional Computing Series.

[Costales et al. 93] Costales, B., Allman, E., Rickert, N. (1993). sendmail. Sebastopol, CA: O'Reilly & Associates Inc.

[Deutsche Welle 95] Deutsche Welle, 1995. http://www-dw.gmd.de/english/index.html

[Deutsche Welle 96] Deutsche Welle: Internet, 1996. Geschäftsbericht 1994, p25, Berlin/ Köln.

[Eisenblätter et al. 93] Eisenblätter, U., Hommes, F., Pless, E. (1993). Ein Vergleich von kommerziellen Netzwerk-Management-Systemen. Arbeitspapiere der GMD Nr. 800.

[gekko 96] http://www.gekko.technopark.gmd.de/about_e.html

[GMD 96] GMD, 1996. http://www.gmd.de/

[Liu et al. 94] Liu, C., Peek, J., Jones, R., Buus B. and Nye, A. (1994). Managing Internet Information Services. Sebastopol, CA: O,Reilly & Associates Inc.

[NCSA 95] NCSA, 1995. http://hoohoo.ncsa.uiuc.edu/docs/Overview.html

[Case et al. 1990] Case, J., Fedor, M., Schoffstall, M., Davin C. (1990). RFC1157 Simple Network Management Protocol (SNMP).

[SATAN 95] Satan 1995. http://www.ensta.fr/internet/unix/sys_admin/satan.html

Author Information

Manfred Bogen has been active in the area of group communication, X.400 development, and X.400 standardisation since 1983. In 1987 he became head of the VaS group being responsible for the provision of value-added services. He studied computer science at the University of Bonn and is co-author of two books about X.400 and distributed group communication. At present, he is the convenor of the TERENA working group on quality management for networking (WG-QMN) and a member of the TERENA Technical Committee and the Internet Society.

Michael Lenz received his masters degree in informatics (computer science) from the university of Bonn. He is working at GMD in the Department for Network Engineering since 1993. His major topics are value-added services, security and information systems. Since late 1993 he has been involved with establishing and maintaining the Web server at GMD and for a German broadcasting company with international program.

Susanne Zier underwent an education to become a mathematic-technical assistant (MTA) at GMD from 1988 to 1991. Afterwards she worked for 2 years in a German hardware manufacturer company. She joined GMD in July 1992 again. At present, she is responsible for the quality management and assurance of the communication servers and Internet applications in different projects with German broadcasters and public administration.

New Pedagogies and Tools for Web Based Calculus

William A. Bogley, Department of Mathematics, Oregon State University, Corvallis, Oregon, USA.
bogley@math.orst.edu. http://www.orst.edu/~bogleyw

Jon Dorbolo, Department of Philosophy, Oregon State University, Corvallis, Oregon, USA.
dorboloj@ucs.orst.edu.

Robert O. Robson, Department of Mathematics, Oregon State University, Corvallis, Oregon, USA.
robby@math.orst.edu. http://www.orst.edu/~robsonr

John A. Sechrest, Department of Computer Science, Oregon State University, Corvallis, Oregon, USA.
sechrest@cs.orst.edu . http://www.peak.org/~sechrest

Abstract: This work is part of a larger project, called *InterQuest*, which is dedicated to establishing effective models and methods for teaching using Internet technology, especially the World Wide Web. Here we discuss features of two InterQuest projects, *CalculusQuest* and *QuestWriter*.

CalculusQuest is a first-term differential Calculus course based entirely on the Web. It was developed around a set of performance objectives and in accordance with the principle that technology should serve pedagogy.

QuestWriter is an environment for administrating and authoring general Web-based courses. QuestWriter includes tools for creating classes and setting up student authentication. It can record, evaluate, and react to the results of on- and off-line activities. It enables instructors to author and administrate *question-based activities* (for example, savable quizzes with feedback) and *communication activities* (pedagogically driven guided student exchanges).

We concentrate on describing those aspects of CalculusQuest and QuestWriter which uses Web technology for specific pedagogic purposes.

Introduction

In the Fall of 1995 at Oregon State University, only 250 of about 500 students initially enrolled in first term Calculus passed the course with a "C" or better [Brumley]. This is typical nationally [Steen 88]. In response to this problem Bogley and Robson joined forces with Dorbolo, who in 1993 had pioneered pedagogic uses of the Internet and Web to address problems with an introductory philosophy course, and Sechrest, who had been providing technological support and vision for Dorbolo's project in conjunction with the educationally oriented Shared Computing Environment which is part of the Network for Engineering and Research in Oregon (NERO).

One result of this collaboration is an ongoing project for the Oregon State System of Higher Education. The goals are to develop a Web-based first term Calculus course, to provide an infrastructure for offering such courses, and to demonstrate a model for teaching Web-based courses across educational sectors. CalculusQuest and QuestWriter are the concrete realizations of this project. They are a team effort which has come to involve content experts, programmers, graphic artists, and an instructional designer.

A beta version of CaclulusQuest will be offered in the Fall of 1996 at Oregon State University, Linn Benton Community College, and to high school students in Eugene, Oregon. It will use many of the tools from QuestWriter, which will be available for platform-specific beta testing this Winter. An extensive program to assess the efficacy of CalculusQuest is in place.

This report is limited to pedagogic applications of technology for which we believe either the pedagogy or the use of technology is new. The details of design methodology, mathematics education issues, the problem of communicating mathematics via the Web, and the architecture of QuestWriter will be addressed elsewhere. The InterQuest [IQ 96] and CalculusQuest [CQ 96] home pages serve as current references.

CalculusQuest

CalculusQuest is broken into 10 *stages*. Each stage has three main areas: a *Lesson* area, a *Practice* area, and an area called *Onward and Upward.* Content is delivered in the Lesson area. The Practice area is the equivalent of homework and labs. Quizzes with feedback are the mainstay of the Practice areas, but some Practice activities are more imaginative. "Onward and Upward" contains activities used to assess progress and comprehension.

We will discuss two aspects of CalculusQuest: (1) How we implement extensive interactivity which engages the student and merges the skills-based and process-based aspects of Calculus and (2) how we use the Web to cater to diverse learning styles and academic backgrounds. Much of the substantial attention paid to Calculus teaching during this last decade [Tucker and Leitzel 1995] [UME Trends 1995] has been focused on these goals.

Interactivity and Engaging the Student

From the very start we engage the student with short, interactive stories. In the first such story the student inputs guesses of the exact height of a fictional character named Andron. Andron gives one of two honest replies: "I am taller than that" or "I am not taller than that". The student discovers that it is possible to estimate Andron's height to any degree of accuracy, but not to pin it down exactly.

The majority of content is presented in a Shaum's outline style. Following accepted principles of Web design [Lynch 96] we have kept pages short and restricted to a single concept. Over half of our pages contain in-line interactive exercises which require thoughtful input. These force students to stop and think about the material and give them the opportunity to assess understanding. Hints and explanations are accessible through separate buttons on the page. These are displayed in pop-up windows. Pop-up windows permit the inclusion of optional information without increasing the length pages or adding another level of links.

In-line interactivity and pop-up windows are programmed client-side using *Javascript* [Javascript 96] [Danesh 96], a proprietary language built into Netscape Navigator versions 2.0 and higher. Microsoft Internet Explorer also implements Javascript but without pop-up windows. A significant use of Javascript is to present the student with "black box" functions for which values can be obtained by entering a number and pressing a button. Numerous activities emphasizing exploratory decisions are centered around black box functions. These activities are unique to the Web and rely heavily on the client-side technology for instantaneous feedback.

Other forms of interactivity are implemented server-side using QuestWriter. QuestWriter generates URL's which are linked from CalculusQuest pages. Examples include self-assessment quizzes and *communications activities*. All exercises in our Practice area give feedback (either through client-side or server-side technology) and can be repeated. Server-side "quizzes" have the added feature that they can be saved for future reference, see the section [Question-based Activities].

Communications activities are interactive exercises which exploit the communications capabilities of the Internet and Web. The most complex communication activities are structured peer-peer exchanges. These are facilitated by our server. For more details see the section [Communications Activities]. A simplistic communication activity is a logout page which demands of the student a short assessment of the learning session. More traditional communication activities include the use of Hypernews, bulletin boards, and chat rooms to foster discussions of Lesson content and Practice area exercises (allowed and encouraged). In their totality communications activities force students to communicate mathematics on a regular basis.

Addressing Diverse Learners

The mathematical community wants texts which are "lean and lively" [Douglas 86]. Client disciplines want texts which impart a large set of "indispensable" skills. Educators insist that we address all learning styles, and administrators demand that we bring along students with diverse mathematical backgrounds. These standards are impossible to meet in a printed text, but the non-linear nature of the Web offers new potential. CalculusQuest strives to meet all of these standards.

Each CaclulusQuest Lesson has a central core of pages accessible from a special page called the Lesson *Hub*. CaclulusQuest is an outcomes-based course designed around a list of performance objectives and criteria, see [CQ 96]. The core pages cover all performance objectives for the Lesson. Diverse interests are addressed by *enrichment pages*, unobtrusively linked and visually distinguished from core pages. These add historical context and present a deeper mathematical point of view. In future versions, students will be able to choose application areas of interest and see pages tailored to this choice.

The inclusion of background material is addressed through a separate *Field Guide to Functions*. The Field Guide offers review of all basic functions used in Calculus. Another project at our university is building pre-Calculus modules which will be linked and coordinated with CalculusQuest. These separate entities, together with their interactive self-assessment opportunities, give students the resources *and responsibility* to recognize and remedy deficiencies.

The organization of CalculusQuest Stages permits students to work through the material in any order. We go so far as to give the students the following table of *typical study approaches*. The approaches model Kolb's learning cycle [Stice 87] and we encourage students to find individual approaches which work best for them.

From Beginning to End	Working Backwards	Back and Forth
I. Read through the Lesson Pages	I. Look at what is in "Onward and Upward"	I. Glance at the Practice Problems and "Onward and Upward"
ii. Try the Practice Problems	ii. Find worked practice problems which might help	ii. Read the Lesson pages a few at a time
iii. Go back to the Lesson Pages as needed	iii. Read the Lesson Pages when you get stuck	iii. After a few Lesson pages, go back to the Practice area
iv. Perfect the Practice Problems	iv. Perfect the Practice Problems	iv. Read Lesson Pages and perfect Practice problems
v. Do the activities in "Onward and Upward"	v. Do the activities in "Onward and Upward"	v. Do the activities in "Onward and Upward"

Figure 1: Typical Study Approaches for CalculusQuest Materials

Students enrolled in Dorbolo's Web-based introductory philosophy course divide themselves into five separate groups based on personal philosophical framework. Each group sees a different set of readings. Future versions of CalculusQuest may divide students into groups with different general goals for learning Calculus.

QuestWriter

QuestWriter is the name given to a collection of course management and course authoring tools which interact with an SQL database containing registration information, authentication information, and a *gradebook*. The gradebook is a collection of *student vectors*. The components of the student vectors are *activities* which can be graded or simply marked as completed. The majority of activities are on-line, but the instructor can create custom activities such as proctored tests. A basic design principle is that information on repeated or re-graded activities is never thrown away. The default is to show only the last recorded score.

The highest level activity in QuestWriter is the creation of a new course. The software creates the directories and database needed to manage the course and to author and edit activities. Students apply for registration via the Web. This process is independent of institutional registration procedures and is filtered through the instructor.

Progress and Grades

The gradebook is a dynamic record of student progress. Students can view their own (and only their own) entries. Instructors can scan the entire class list. This is a valuable aid to student retention. A conscientious instructor can literally check for stragglers each morning and, since all students are on-line, send encouraging or inquisitory email messages to students in danger of falling too far behind. The key is the timeliness of the information and the immediate contact with the students.

The instructor can weight and curve activities and thus generate a score for the class, but QuestWriter also has the ability to archive and retrieve detailed components of a student's performance rather than just the final grade. This raises new possibilities. By adjusting weights, for example, different departments might use different criteria to determine if the same student taking the same course has met program requirements. The QuestWriter gradebook will eventually allow us to build pages which deliver material *dependent upon a student's current state of progress or advancement*. Finally, QuestWriter is a prototypical of the administrative tools needed for Oregon's standards-based K-12 educational reform (House Bill 3565) and for the Oregon State System of Higher Education's new Proficiency-based Admissions Standards System [PASS] [Conley & Tell 1996].

Question-based Activities

Our original intent was to build an environment for authoring quizzes that offered optional feedback and drew questions randomly from a database of questions. We have not yet implemented the randomization procedure, and we have come to realize that "quizzes" represent a general form of interaction as characterized in the following figure. We call these *question-based activities*. Answers to question-based activities can be recorded for future reference and/or evaluated either for the gradebook or just to provide feedback.

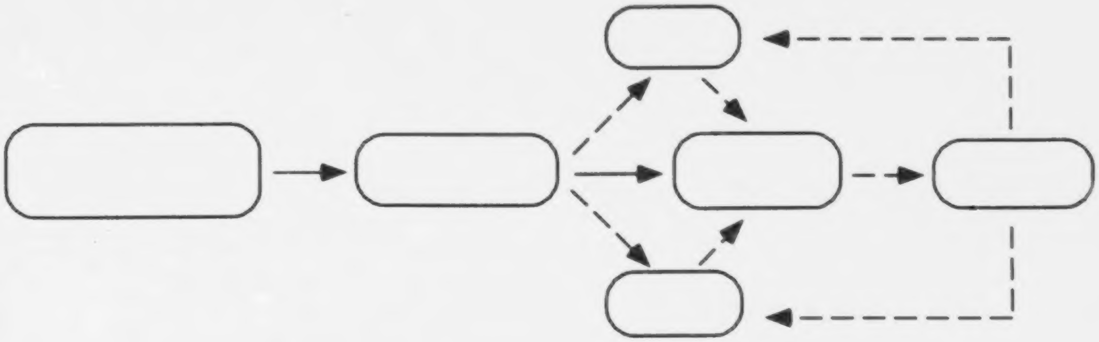


Figure 2: Feedback Loop for Question-based Activities

Figure 2: Feedback Loop for Question-based Activities

This model can be used for graded on-line quizzes and tests, for self-assessment exercises where the emphasis is on feedback, and for building student notecards. A *notecard* is a Web page containing HTML forms with questions for the student to answer. After correction a version of the page incprporating the student answers can be saved as a permanent and personalized reference. A 1994 site visit review of Web-based philosophy materials generated the comment that, at the end of the course, the students had nothing akin to a text or set of notes to take with them. Notecards respond to this: a virtual rolodex of notecards can guide a student through a Web-based or partially Web-based course and provide a permanent record.

Question-based activities can also be used in conjunction with the gradebook as gateways which block access to material until a certain level of response is attained. This has wide application as an internal pedagogic tool and as an external placement device.

QuestWriter offers an environment for authoring question-based activities linked to particular classes. Currently, we support three types of questions ñ true/false, multiple choice, and fill-in-the-blank. The instructor constructs a header and footer for each activity, generally containing instructions and links to URL's which the student might want handy while completing the activity. Each response to each question has a place for optional feedback. Different types of questions may be included in one activity and questions may be edited, added, and deleted from existing activities The instructor may choose between various evaluation options and may indicate if the activity can be saved by the student.

Communications Activities

Students in Dorbolo's Web-based philosophy course are required to engage in "discussion activities", now termed *communication activities*. A typical such activity is a simple exchange between two students Alice and Bob. Alice and Bob receive instructions to write a letter to Plato. The letters are exchanged and each student, in the persona of Plato, replies to the other. Other activities involve loops of three or four students.

Communication activities require special automated administration. The fundamental problems are forming groups, passing information among students in a group, and insuring that progress is being made. QuestWriter contains an authoring tool for communications activities which involve 2 students and any number of "passes". Thus Alice could send off an assignment (one pass), receive input (2 passes), send off a reaction (3 passes) and receive a reaction (4 passes).

If there are *n* passes, then there are *n* roles to be played in the communication chain. *In our model, each student plays each of the roles.* This raises the question of how to pair up students. We have two ways.

The exchange model: As students declare themselves ready, they are paired off and stay together. Initially, both partners complete the first pass and exchange results.

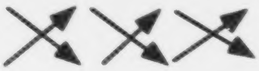


Figure 3: The Exchange Model of a Communication Activity

The "lazy scheduling" model: Say Alice is ready to begin a discussion activity. She receives instructions and sends off the first pass. This goes to the first student (other than Alice) who is ready for the second pass. This could be Bob. Alice is ready for the second pass, but her input could come from *any* student, not necessarily Bob.

BEST COPY AVAILABLE

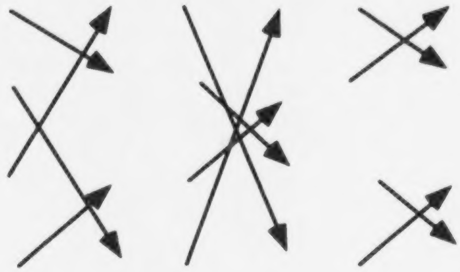


Figure 4: The "Lazy Scheduling" Model of a Communication Activity

The assignment of partners is facilitated by the gradebook. Deadlines can be set for each pass and for the entire activity. QuestWriter generates automatic email reminders and, depending on the model used, will reassign students if a partner is not participating. The instructional staff can masquerade as students. This is necessary to deal with a "left over" student and provides a second monitoring mechanism.

QuestWrtier includes an instructor front-end for both the authoring, viewing, and evaluating communications activities. The instructor can see all responses made by a particular student or follow a thread. There is also a viewing tool for the students.

Summary

The Web and Internet comprise a new medium with new pedagogic opportunities and challenges. CalculusQuest takes advantage of client-side and server-side technology to implement extensive interactivity aimed at provoking student thought while making the learning experience enjoyable. Artifices such as varied background colors, hub pages, and pop-up windows are used to create a content-rich environment which is none-the-less easily negotiated and flexible enough to accommodate diverse learning styles and backgrounds.

Creating and administering a Web course is a time-consuming and daunting prospect. QuestWriter provides the infrastructure without which this cannot happen on more than an ad hoc basis. The SQL database at the core of QuestWriter has great potential in the contexts of outcomes-based and Web-based education. We have introduced two general concepts, the question-based activity and the communication activity, which use the communications potential of the Internet and Web for a variety of specific pedagogic purposes.

Concrete issues we have not addressed include the problems and frustrations of dealing with emerging technology and the difficulties of communicating mathematics on the Web. We have not detailed the actual team process of building a course like CalculusQuest and have only touched upon some fascinating issues of educational Web page design. We have also not discussed many technical points of CalculusQuest and QuestWriter, nor have we said anything about issues of privacy, data protection, and authentication. We intend to report on all of these things in appropriate forums. Hopefully by the time we do so we will have hard data speaking to one overriding question: *Did it work in practice?*

Acknowledgments

InterQuest is funded by the Oregon State System of Higher Education, Technology Resource Fees and the Instructional Media Committee at Oregon State University, and by Linn Benton Community College. Funding for NERO is provided by NASA. Our content development team was augmented by Judy deSzoeke (Linn Benton community College) and substantial contributions to CaclulusQuest and QuestWriter were made by Eric Altendorf, Matthew Hall, Jill Lombaer, Jeff Ronne, and Srikanth Uppala, all students at Oregon State University.

References

[UME Trends, 95] (1995) Special issue devoted to the history and assessment of Calculus reform. UME Trends 6 (6), 1-32

[Tucker & Leitzel 95] Tucker, A. & Leitzel, J.R.C, editors. (1995). Assessing Calculus Reform Efforts. Mathematical Association of America, Washington, DC

[Danesh 96] Danesh. Teach Yourself Javasxcript in a Week. (1996). SAMS

[Douglas 86] Douglas, R., editor. (1986) Toward a Lean and Lively Calculus. MAA Notes; no. 6, Mathematical Association of America, Washington, DC

[Steen 88] Seen, L., editor (1988) Calculus for a new Century. MAA Notes; no. 8, f Mathematical Association of America, Washington, DC

[Lynch 96] Lynch, P.J. Web Style Manual. (1996) Yale Center for Advanced Instructional Media. http://info.med.yale.edu/caim/StyleManual_Top.HTML, Section II.

[Stice 87] Stice, J.E. Using Kolb's Learning Cycle to Improve Student Learning, Engineering Education, February, 1987.

[Javascript 96] The Javascript Authoring Guide. (1996)

http://home.netscape.com/eng/mozilla/Gold/handbook/javascript/index.html

[Brumley 96] Statistics generated by Carolyn Brumley using Oregon State University's Data Warehouse interface to the campus student database. (1996)

[Conley & Tell 96] Conley, D. & Tell, C. (1996). Proficiency-based Admissions Standards System Project: Frequently Asked Questions (and separately) Admission Standards: Content & Process Areas, Proficiencies, and Indicators. Oregon State System of Higher Education, Eugene, OR.

[PASS 96] Proficiency-based Admissions Standards System Home Page. http://pass-osshe.uoregon.edu

[IQ 96] InterQuest Home Page. (1996) http://www.sce.ojgse.edu/iq

[CQ 96] CalculusQuest Home Page. (1996) http://artemis.orst.edu/CQ

Supporting Teaching and Learning Via the Web: Transforming Hard-Copy Linear Mindsets into Web-Flexible Creative Thinking

Ellen Yu Borkowski, Teaching Technologies, ellen_borkowski@umail.umd.edu
David Henry, Academic Software Integration and Development, david_henry@umail.umd.edu
Lida L. Larsen, inforM/On-line Information Resources, lida_larsen@umail.umd.edu
Deborah Mateik, Electronic Information Services, deborah_mateik@umail.umd.edu
Academic Information Technology Services, University of Maryland,
College Park, Maryland, USA

Abstract: This paper describes a four-tiered approach to supporting University of Maryland faculty in the development of instructional materials to be delivered via the World Wide Web. It has been implemented leveraging existing equipment and staff by the design of Web posting, editing, and management tools for use on the campus-wide information server, inforM, and training the trainer instructional modules. Key faculty and staff are identified and assessment of skills and needs are conducted. Individualized support packages are designed using part, or all, of the four tiers: data entry, scanning, converting and formatting of documents, short course training in the use of the Web and, specifically, of HTML, forms-driven posting and editing tools, and electronic publication and evaluation skills; faculty-focused training and mentoring in instructional technology and pedagogy; and integrated use of cutting edge teaching theaters with continuous instructional technology staff support.

Background

The University of Maryland at College Park is a Land Grant University that serves 35,000 students and 11,000 faculty and staff. It is composed of 13 Colleges and Schools and 120 academic departments. Information technology requirements and capabilities vary widely. Some faculty have access to the cutting edge resources of electronic classrooms, others have only passing acquaintance with student-oriented open workstation labs and e-mail. The World Wide Web is fast becoming a technological equalizer between these two extremes. Web-based education is possible on this campus thanks to a pervasive network infrastructure (every classroom and office on campus is wired) and a well-developed central Website: inforM (http://www.inform.umd.edu). Faculty and staff may create a Website for themselves, their courses, their research field, or their departments on the system. Academic Information Technology Services (aITs) and its inforM staff--which consists of one content coordinator and 1.4 systems administrators, plus a group of talented student staffers--provide coordination and training for information owners. Our ability to provide valuable and timely information via the Web is directly related to our ability to identify committed information owners, and to provide them with good tools to use and training and support for those tools.

Identifying potential information owners is relatively easy. Engaging some of them is easy--they are already knocking on our door. Engaging others often is not. Even if they understand the importance of their contribution and want to be "on the team," they often believe they lack the time to create and maintain information in a web-format. We have identified the key problem elements as insufficient knowledge, time, experience, and/or desire to transform this information for Web use. A set of Web tools has been developed in response to all of these elements except the necessary desire to be "on the Web". This is not our concern.

It will be noted that human resources are a critical element in both the training of our faculty and in the faculty themselves. When new technologies emerge they must be used appropriately to supplement or supplant current instructional technologies. We do not enjoy adequate staffing to provide multiple formats of instructional materials. Choices must be made when determining the best format to be used in any setting. Faculty must ask themselves if their Web pages will be the primary source of course materials from which other formats are produced or vice-versa.

Identifying and Meeting Needs of Information Owners with Differing Skills

As with any large institution, our potential information owners have a wide range of computer and Internet skills. Our information owner tools:

- Web Spinner (posting/editing forms driven tool located),
- Calendar Submit (submittal tool for on-line calendar),
- Syllabus Submit (a forms driven tool for posting course information),
- Expire (allows information owners to preset expiration dates for files),

and others have been designed to the low-end of the technical skill range in order to make them the most useful to the largest percentage of the campus population. Like most tools, e.g., a sewing machine or hammer and chisel, they each can be used to create something basic or something complex: a sleeveless vest or an evening gown. The key to the finished product is often in the training of, and expertise in the use of, the tool, and, of course, in the raw materials available for use. We have chosen a four-tier training and support approach to create that base of expertise for faculty information owners on our campus.

Tier One: We Do It All For You

Tier One information owners tend to be those who understand the importance of including their course information on the Webserver, but who lack the expertise, time or interest to process the materials themselves. For these users, we provide a student assistant to process both electronic and hard-copy documents and post them to the system in text format. A scanner is provided for hand-written class notes and graphic materials. Though graphics and HTML support is available, it is not standard at the Tier One level. In certain politically expedient cases, text files are converted to HTML and enhanced with graphics to illustrate to the Tier One information owner the benefits to becoming a Tier Two information owner. It is done frequently when assisting a department to establish a Web presence. A student staff member will work with the departmental liaison to create the initial Web pages while simultaneously training that person in the use of our information owner tools so that they can manage and expand the pages without further assistance in the future.

Moving Tier One Information Owners to Tier Two

The information owners at the Tier One level are strongly encouraged to move to the Tier Two level by taking standard training offered by aITs to learn more about the Internet, inforM and our web management tools. Often, specialized Web in-service training is arranged for an entire department. These sessions focus on potential uses of the Web for their users (e.g., students and colleagues). For example, a session for the faculty of Women's Studies would demonstrate the use of the Web to search inforM's internationally renowned Women's Studies database for research and curriculum development purposes. It would also show the ease with which faculty could make additional contributions to that database to add their expertise to the evolving knowledge base. The session would also demonstrate innovative Web pages created by other faculty members for course materials and specific instructional units, including those utilizing cgi-bin programming, forms support, and other Web technologies.

This effort includes frequent marketing of available Web tools, techniques, and services in regular and guest columns for specialized university publications; active recruitment of specific departmental and research center liaisons; and, high visibility of inforM staff in campus-wide projects and programs. Tier One information owners often feel compelled to move to Tier Two when they understand the full potential of the Web to support them in their teaching and research endeavors. The ability to manage their own Web pages is also a key factor in their decisions to move to the Tier Two level.

Tier Two: Swinging on the Web with a Safety Net

Tier Two provides on-line tools and tutorials with both centralized and decentralized consulting support. All of our basic tools are available on-line. They are password protected if they are input programs such as Web Spinner and Calendar Submit. Student assistants in our office and faculty/staff liaisons in most colleges and some departments have been trained in the use of these tools and can provide help by telephone or in person. Passwords are given to authorized information owners who agree to abide by inforM Guidelines. They are asked to review our on-line tutorials and guides. They are given the name and contact information for our staff and for the contact person in their college or department.

The Standard Scenario

A standard scenario begins when an information owner agrees to manage a specific directory of information on the system, be it a directory of files supportive of a specific course or a database of discipline-specific resources. Authorization is verified with the individual's college or department. Then training needs assessment begins when the inforM Coordinator meets with and assesses the information owner's access to appropriate technical infrastructure and her/his organizational and technical skill levels. The inforM Coordinator then decides with her/him the general formats and input tools that will be used. The Web Spinner tool is demonstrated and the tutorials and guidelines (http://www.inform.umd.edu/CompRes/WWWRes) are pointed out. Training needs are identified and group training opportunities are recommended accordingly. Quite frequently self-selecting faculty operate at a higher technological level and are more "net"-experienced than most. They do well with a minimal amount of training and freedom to independently experiment with the tools provided.

For those faculty requiring a bit more support and training, aITs staff provide regularly scheduled "short courses" on such topics as HTML, Web browsing, and related Internet technologies throughout each semester during working hours. Additional training is available in similar courses taught by University students during non-working hours. These classes are a big draw for students of faculty who have made the commitment to delivering class materials via network technologies, like the Web.

The Web Spinner tool allows information owners to post text or other files in a gopher type tree structure or to create a standard Website of one or more HTML pages. Information owners can freely choose either method. When using the gopher type tree structure, Web Spinner provides "on-the-fly" conversion of the information into an HTML default format. When creating a standard Website, the information owner hard codes or uses the HTML editor of his/her choice and posts the homepage as "index.html".

New information owners are encouraged to build their files on their local hard drive using a package with which they are already comfortable and then to post them using the Web Spinner tool. This can be done by "copying" the file and then "pasting" it into place. Web Spinner can then be used to access files for on-line editing. Web Spinner options include:

- traversing directories and
- moving files and directories
- entering current files to edit
- adding URLs
- making directories
- changing passwords
- posting files
- getting help
- deleting files and directories

Some standard cgi-bin programs and Java scripts have been created for the use of individual information owners. They can do a "copy/paste" of the script, inserting their own text and/or graphics, and including the script within their HTML page.

To support the instructional use of the Web technology beyond the inforM server we offer access to a second Webserver with fewer security controls. This is in response to faculty who wish to do their own cgi-bin or other programming, or want to provide "webspace" for students to work with cgi-bin or other programming, and/or to post student class projects. We have created a virtual "Office of Technology Assisted Learning" (OTAL) which is comprised of lead staff from inforM and Teaching Technologies, another key unit in aITs. Faculty may submit their own proposals or "sponsor" student proposals for use of the OTAL Webserver. Most projects have a one semester duration. No technical consulting support is provided.

Forms-Driven Templates

Two new forms-driven templates, one for course information and one for faculty/staff information, are being developed at this time that will bridge Tier One and Tier Two. These on-line forms will allow faculty or their support staff to place basic information on inforM for themselves and/or their courses without requiring any involvement of the inforM staff. The forms will be verified using campus IDs and associated PIN information and, for the course information, the semester's current database of course instructors and teaching assistants. The information will be posted in a standardized format and will be linked to from the on-line Schedule of Classes.

Tiers One and Two serve the majority of the information owners on the College Park campus. No UNIX or advanced skills are necessary to create a Website. Basic information is provided to our users. General guidelines for format and content are followed. A small fraction of these information owners have spent the time necessary to develop advanced HTML skills and have creatively used our disk space to implement highly effective Websites. In cooperation with these individuals, the campus' Center for Teaching Excellence and the Teaching Technologies office, Tiers Three and Four have been developed.

Tier Three: Training Web Innovators--Institute for Instructional Technology

Faculty are the primary engineers of curricular development on the University of Maryland at College Park campus. However, many faculty are daunted and confused by the vast array of computer, information, and communication technologies at their disposal. For effective use of campus technology, faculty need support in terms of training, resources, and feedback.

Faculty-Focused Training and Mentoring

Tier Three targets faculty and their technical support staff who have an interest in developing advanced HTML and instructional technology skills to create course materials on-line. In response to faculty support requirements, the Institute for Instructional Technology (IIT) was developed as a collaborative effort between aITs and the Center for Teaching Excellence, a unit that promotes faculty innovation and peer mentoring to improve the quality of student learning on campus. Institute agenda (http://www.inform.umd.edu/TeachTech/IIT/) are comprised of theoretical and practical information taught by pedagogical experts, academic information technology staff, and faculty mentors. Modules include:

- Pedagogical Frameworks
- Graphic Design
- Instructional Use of Listservs/Mail Reflectors
- Legal Issues
- Electronic Publication
- Web Spinner
- Website Evaluation Techniques
- HTML

Each Institute takes up to 20 participants with little or no experience and trains them to be key users and innovators of the Web technology. To participate in IIT seminars and workshops, interested faculty submit proposals outlining their goals for transforming an existing curriculum through technology; their acceptance into the program is based upon the appropriateness of their proposals to the actual objectives of the workshops and upon the space available.

While the IIT provides concentrations in such areas as electronic communication, graphic design and multimedia courseware development, and World Wide Web-based instructional delivery systems, by far the most popular has been the Web module. In fact, some of the most innovative products developed by IIT "graduates" have been Web pages supportive of course curriculum.

On the first day, participants were taught the basics of HTML programming, explored some examples of web-based syllabi and other classroom support materials, and were mentored through a two-hour workshop period (where the ratio of lab assistants to participants was approximately one to three). It is interesting to note that, while most faculty plunged through the challenge of learning simple HTML coding, it was the collective opinion of the class that in the long-term it is more important for faculty to understand the possibilities and limitations of Web page development, and to find a technical support person or talented student to perform the coding tasks. Most faculty see themselves as instructional designers, not technicians.

The focus shifted from text to graphics on day two with a brief introduction to scanning and Adobe Photoshop. Participants were given guidelines for techniques for effective electronic publishing (e.g., usable image sizes and placement options, screen composition, etc.). Another afternoon was spent in workshop format continuing the construction of faculty pages. The high ratio of helpers to participants was useful on this day in particular because, surprisingly, most faculty had little or no experience with file transfer protocols and required close handholding when it came time to transfer their graphics from the desktop to their pages. On the final day, Web Spinner instruction, fine tuning Web pages and managing their own file structures on the Web were covered. Participants also had a peer mentoring session with a faculty member who has successfully integrated the Web technology into the conduct of her classes. The mentor provided no hands-on training, but rather gave IIT faculty a pedagogical overview of the process. The mentor candidly shared insights into how the infusion of this new technology into her curriculum has altered the way she teaches those classes and how it has changed the ways her students learn in class.

What a Commitment to Web Technology in the Classroom Means to Faculty and Students

Discussion arising from the mentored session in this first IIT explored such issues as,

- "is the Web the great panacea for all courses?" (e.g., nothing is gained by introducing technology simply for technology's sake);
- "do most students have the basic technological background to make use of the Web?," and "is there sufficient access to the technology on campus for all students?";
- "is there a recognized 'pay off' for faculty who publish on the Web?" (therefore making it more worthy of their time and consideration); and
- "is there a way to evaluate the impact that technology, and specifically the Web, has on student learning and retention of information?"

Not all of these questions were definitively answered during the peer mentored session. Most participants, who represented disciplines as diverse as Landscape Architecture, Geography, Education, Journalism, and Engineering, believed the Web offers an information delivery medium appropriate to the focus of their coursework. In fact, the proposals they submitted in order to enroll in the IIT were reviewed specifically to ensure that Web technology is, indeed, an appropriate delivery device for the content of their curriculum.

The recognition of the World Wide Web as a valid publication medium for tenure-track faculty is a battle being fought on most campuses around the world. At the University of Maryland, faculty are encouraged to mention electronic publications, Web-based curriculum development, and participation in such enterprises as the Institute for Instructional Technology on their vitae. However, at this time, electronic publication is not weighted equally with more traditional publication genres.

The evaluation of the effectiveness of the integration of this technology into the curriculum will come about as more faculty make use of it and report their experiences to their peers and administration. One of the faculty mentors for an IIT module was able to informally evaluate her Web experience by offering two sections of the course: one taught through traditional lecture and report means, and the other taught by integrating the World Wide Web into all facets of the coursework. In the traditional class, all reading materials were distributed or pointed to in libraries; students wrote reports with word processors and gave in-class reviews of their research. Collaboration on projects was sometimes hindered by incompatible schedules and the like. Students in the Web-based section were directed to readings on-line and to library resources. Most students also used the Web to search out additional reading materials beyond the assigned information (because it was so easy to do). All student research and reports were put together and delivered on the Web. Students were able to collaborate at least partially electronically and continued to make changes to improve the pages even after a grade had been assigned to a project. For both classes, the average grade was a B (project grades were higher in the Web- based class). The faculty mentor indicated that the most telling evaluative feedback came from the students themselves. While students in the Web-based class felt there was more time invested in projects and a steeper learning curve at the start of the class, their sense of accomplishment was greater, the collaborative process was more fruitful, and their confidence in their abilities to do to more integrated, independent research was sharply increased.

What We Learned About Coaching Web Spinning Faculty

Faculty feedback from the first Institute for Instructional Technology module served to refine the direction of faculty Web page support in two areas: the structure and content of future Web Page Development training events and adjunct Web page support services for faculty.

When IIT sponsored its next Web Page Development module several changes were made to the curriculum. A fourth day was added to provide training time in the use of FTP, to provide additional training in a few advanced HTML concepts (e.g., backgrounds, tables), to expand the mentored hands-on workshop periods, and to make time for peer critiquing of the pages under development. A fifth day was added in which standards for citing electronic resources were offered and Website evaluation techniques were covered. In the Website evaluation techniques session, participants were taught to look for the elements of scope, authority, bias, accuracy, timeliness, permanence, added value, and presentation. This enables them to more objectively review their own Web pages for these elements.

Tier Four: Teaching Technologies

Tier Four targets a specific population of faculty who are teaching in the campus' electronic classrooms or who have requested instructional technology assistance for their classes. The electronic classrooms, called Teaching Theaters (http://www.inform.umd.edu/TT/), are advanced electronic classrooms with well-integrated hardware, software, and networking. Software tools have been developed to enhance the collaborative learning environment available in these classrooms. With the introduction of the Web, additional capabilities now exist for faculty using these facilities to support collaboration both in and out of class. The Teaching Technologies office (http://www.inform.umd.edu/TeachTech/), which manages these facilities, works individually with the faculty member to develop his/her course materials and student projects utilizing the campus' inforM server and a Teaching Technologies webserver. The Teaching Technologies Webserver provides interactive Web capabilities (including cgi scripting) and supports individual student projects. Student projects are created individually and/or collaboratively depending on the particular course.

Web pages are provided for all the courses that are held in the Teaching Theaters that link to the course description and syllabus plus the instructor's home page, if available. Another feature added to the pages this current semester is WebChat(tm) discussion pages. WebChat(tm) is a real-time fully multimedia chatting application for the Web. Users can quickly incorporate images, video and audio clips, and "hotlinks" into their chat. This tool has supported discussions of topics both within and outside of class. It has also allowed the support of a collaborative project between an Art History course being taught in the Teaching Theater and an Art Studio course being taught in a traditional classroom. Art History students were required to write a textual analysis of an image which was then given to their Art Studio team members. The Art Studio students then had to create the image from the textual analysis without having seen the original image. Discussion about the process and final results were conducted over the Web using WebChat(tm).

Most of the faculty using the Web support provided by Teaching Technologies have gone through the intensive week(s) of the Institute for Instructional Technology mentioned above. Support for developing materials is provided prior to the semester they are going to utilize the materials. Each faculty member has the opportunity to use both instructional experts and computing experts to develop courses taught with integral use of the Web and other instructional technology on a daily basis.

A vital part of Teaching Technologies is its support team. The team includes both technical and pedagogical experts. At least one support person is present during every use of every electronic classroom. Instructional support is provided to the faculty member prior to and during each semester to assist in preparation through group and individual training. Additional group sessions are organized for the exchange of ideas and experiences. These classrooms offer unique research opportunities in such areas as the study of classroom behavior, teaching strategies, comparative methodologies, and effectiveness of technology.

Conclusion

We feel that our current four-tier approach to meeting the support and training needs of our faculty information owners is appropriate for the variety of technical capabilities and inquisitiveness demonstrated by that group. As our base of faculty Web Spinners grows in the coming year, we will continue to develop more tools and training venues that will facilitate the development of course materials and integrated, discipline-specific databases, while leveraging modest staff resources that are not scheduled to grow in that same timeframe. If these tools and training truly do their jobs, then faculty can, indeed, be the primary architects of academic Web content. They can concentrate on transforming traditional linear forms of delivering information to their students and colleagues into creative and collaborative associations of information via the Web without becoming mired in code and procedures. The fifth and final tier of user support: self-support and full ownership of information that advances contributions to the curriculum and discipline on the part of College Park faculty.

The WWW as a Primary Source of Customer Support

Alexandra Bost, Webmaster
SAS Institute Inc., Cary NC USA
alex@unx.sas.com

Abstract: First released in October 1994, SAS Institute's World Wide Web (WWW) site was one of the first corporate sites of its kind available on the Internet. The site quickly became more than just a useful "message;" it became a significant source of user support. Observing this trend, SAS has implemented a variety of user assistance methods: Online Order Systems for SAS Documentation; Online Registration for all Training Courses; and a Technical Support FAQ that contains over 1000 of the most frequently asked questions on our Technical Support telephone lines. These resources are updated more quickly and efficiently than their "paper "alternatives.

Introduction

First released in October 1994, SAS Institute's World Wide Web (WWW) site was one of the first corporate sites of its kind available on the Internet. The main objective of the SAS WWW has always been twofold: first, to provide information about Software and Business Solutions provided by SAS Institute; second, to provide Online Services and Resources to our world-wide customer base.

By the time the WWW site was a year old, it became apparent that our useful "corporate message" tool had evolved into an effective means of consolidating service and support resources. While many visitors to the site were reading the Sales and Marketing information on the first few pages, many more were visiting the Technical Support, Publications, and Training pages.

Observing this trend in the need for User Support, we have implemented a variety of user assistance methods in our Customer Services area of the web site: Online Order Systems for SAS Documentation; Online Registration for all Training Courses; and a Technical Support FAQ that contains over 1000 of the most frequently asked questions on our Technical Support phone lines. These resources are updated more often than their "paper" alternatives. This paper will demonstrate how we use the WWW to support a variety a of customer needs, more quickly and efficiently than traditional methods.

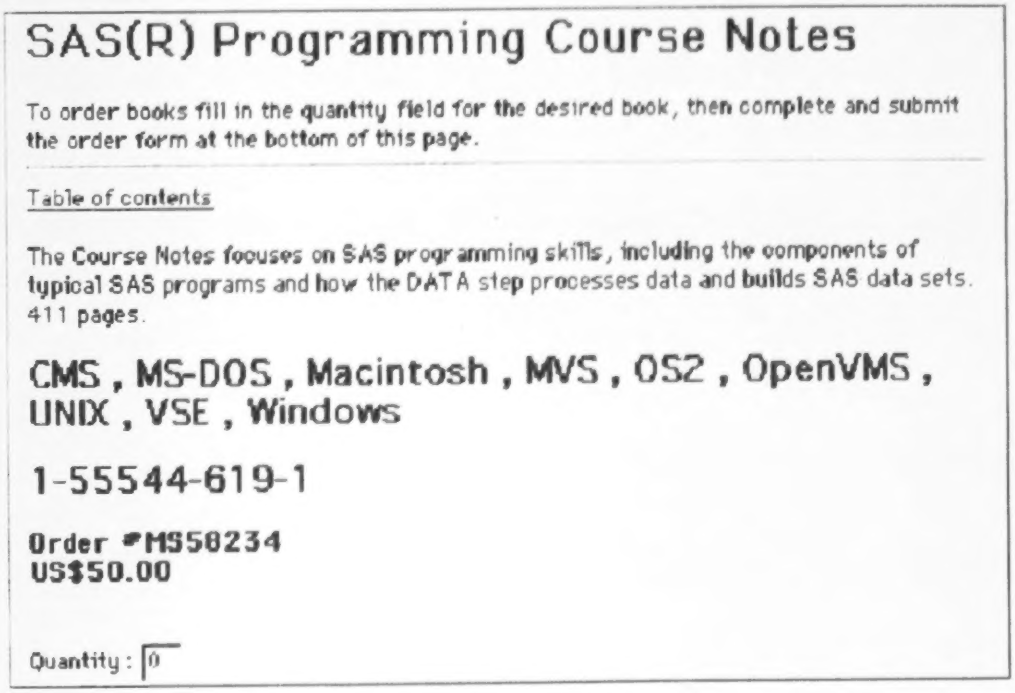
The SAS Institute Web site, now composed of nearly 20,000 pages of detailed information, averages nearly 1,500 visitors each business day. Of these visitors, almost 1100 are customers seeking information, resources, and solutions to problems.

The Customer Services home page, available directly from the SAS WWW home page, provides a wide spectrum of online resources. Three of the most popular are: An online Publications Catalog and ordering system; Training Course Descriptions and Registration; The Online Technical Support FAQ.

The Online Publications Catalog

The Online Publications Catalog contains the same information available in our print version of the catalog; however, with the addition of a searchable index and online order system, our customers are able to take advantage of the most current information. While the print version comes out twice a year (and quickly becomes obsolete), the online versions are updated weekly. In addition, three versions of this catalog are available from our web site: The United States version, the Canadian version, and the Japanese version. The Online Publications Catalog is available via the Documentation/Publications section of the Customer Services home page.

Each version of the Publications Catalog contains over 400 items; therefore, the print version of the catalogs is large, and the topics are not always easy to find. With the WWW version, users no longer have to place a phone call to find a title relating to a product; they can easily and efficiently browse a database that is sorted by title, product, or even new releases [Fig. 1]. Another advantage of the Online version of the catalog is the link to a sample "Table of Contents;" this option allows users to preview the contents of the book.



SAS(R) Programming Course Notes

To order books fill in the quantity field for the desired book, then complete and submit the order form at the bottom of this page.

Table of contents

The Course Notes focuses on SAS programming skills, including the components of typical SAS programs and how the DATA step processes data and builds SAS data sets. 411 pages.

CMS , MS-DOS , Macintosh , MVS , OS2 , OpenVMS , UNIX , VSE , Windows

1-55544-619-1

Order #MS58234
US$50.00

Quantity : 0

Figure 1: Example Publication from Database, complete with Order Form.

To create the searchable database, a WAIS index is built against the hundreds of entries in a SAS data set. Once the index is built, a perl script queries the data and extracts the index-building topics. Once the index is complete, the HTML page is created as a front-end to the index:

Sample Code - (partial Perl script)

```
$waisq = "/dept/pub/web/servers/bin/waisq";

$waisd = "/dept/www/test/wais";

$src = "pubcat";

$title = "Publications Catalog";

sub send_index {

print "Content-type: text/html\n\n";

print "<HEAD>\n<TITLE>Index of ", $title, "</TITLE>\n</HEAD>\n";

print "<BODY>\n<H1>", $title, "</H1>\n";

print "This is an index of the Publications Catalog on this server.\n";
print "Please type a query in the search dialog.\n<P>";
print "You may use compound searches, such as: <CODE>MVS AND VM</CODE>\n";
print "<ISINDEX>";
}
```

Processing the Order

Once users have found the book(s) they would like to purchase, they indicate the quantity needed, and submit an order form at the bottom of the current page. The order is then e-mailed to an account that handles the sale. The order is confirmed within minutes (via e-mail) or later that day (via a phone call) and the books are shipped immediately. In contrast, the print version of the catalog requires a phone call, fax, or surface mail to place an order; the process can take up to a week longer.

All information on publications is updated daily by the SAS Publication's staff and stored in a SAS data set; the information from the SAS data set is exported into HTML, and this is then copied into the web server. The WAIS index is rebuilt weekly, or more often as needed.

Future enhancements to the Online Publications Catalog will include allowing transaction payment via credit card through a secure web server.

Training Course Descriptions and Registration

The SAS Training Services section, available via the Training section of the Customer Services home page, also provides a US and Japanese version of the WWW pages. From this section, a user can browse a list of over 100 training courses. Information is available on training locations, training software, video based training, and descriptions of Public Courses.

Use of the online Training Services area provides a timely and thorough way for all SAS users to keep up with course schedules. In North America alone, there are over 100 courses available at more than 22 locations; the importance of updating both the courses and location information on a weekly basis cannot be stressed enough. While the print version of the training guide is updated quarterly (and quickly becomes obsolete), the Web version is updated weekly. The course information is stored in a SAS data set and updated by SAS Training staff; then, once a week, a SAS program is run that converts the SAS data into HTML:

Sample Code - (partial Perl script)

```
libname titles 'edc.cdev.saslib' server=mvs.pr6xhsrv;
libname courses 'edc.cdev.courses' server=mvs.pr6xhsrv;
filename out "&outdir/bogus.file";
options nofmterr;
/* read in the titles data set */
data titles; set titles.books(keep=ccode title status type key product1-7;
length word $ 40 temp $ 200;
if status = 'X' and type = 'A';
n=1;
```

BEST COPY AVAILABLE

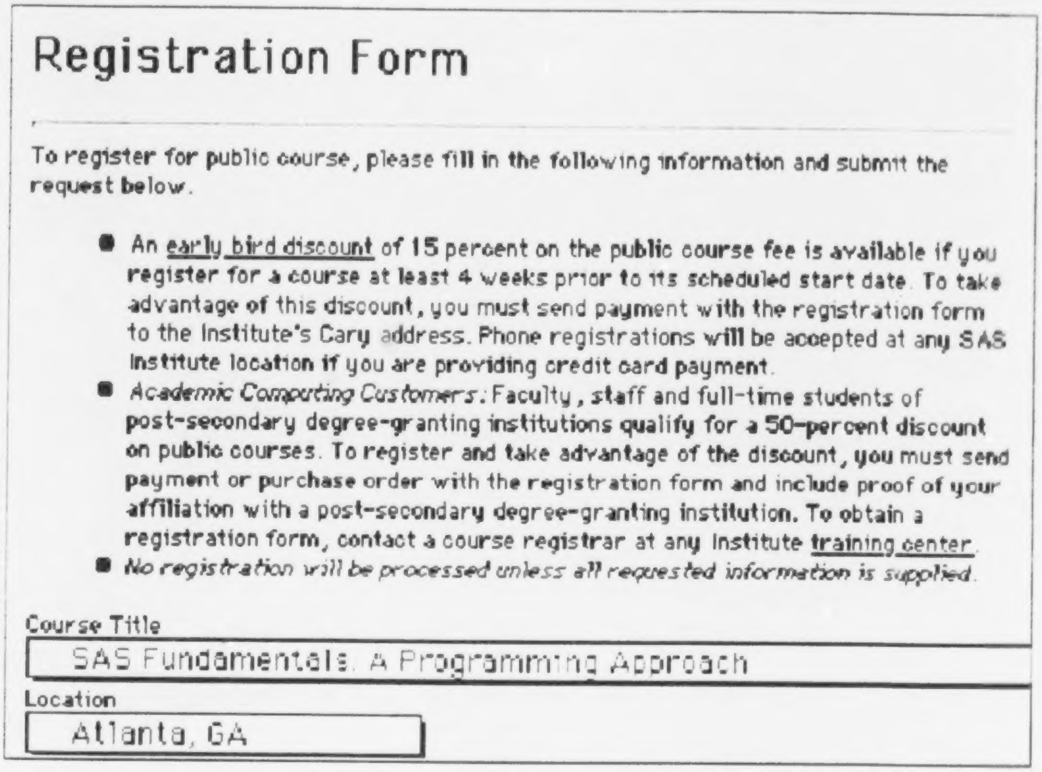
```
do until (word='');

word=scan(title,n,' ');

if trim(left(word)) = '(R)' then do;

n + 1;

word=scan(title,n,' ');

end;

if index(word,'(R)') then

word = substr(word,1,index(word,'(R)')-1)||

substr(word,index(word,'(R)')+3);

if n = 1 then

temp = trim(left(word));

else

temp = trim(left(temp))||' '||trim(left(word));

n + 1;

end;

title = temp;

drop word n temp status type;

run;
```

Registering Online

With courses constantly being added or updated, this is the most accurate way for a customer to be sure they are receiving timely information. Once a user has made a training choice, they can submit an online form to an account in the Professional Services Department that pre-registers them for a course [Fig. 2]. In addition, a user can simply request more information.

Future enhancements to the Online Course Registration Form will include full course registration, with credit-card payment allowed through a secure web server.

BEST COPY AVAILABLE



Registration Form

To register for public course, please fill in the following information and submit the request below.

- An early bird discount of 15 percent on the public course fee is available if you register for a course at least 4 weeks prior to its scheduled start date. To take advantage of this discount, you must send payment with the registration form to the Institute's Cary address. Phone registrations will be accepted at any SAS Institute location if you are providing credit card payment.
- *Academic Computing Customers:* Faculty, staff and full-time students of post-secondary degree-granting institutions qualify for a 50-percent discount on public courses. To register and take advantage of the discount, you must send payment or purchase order with the registration form and include proof of your affiliation with a post-secondary degree-granting institution. To obtain a registration form, contact a course registrar at any Institute training center.
- *No registration will be processed unless all requested information is supplied.*

Course Title

SAS Fundamentals: A Programming Approach

Location

Atlanta, GA

Figure 2: The Online Course Registration Form

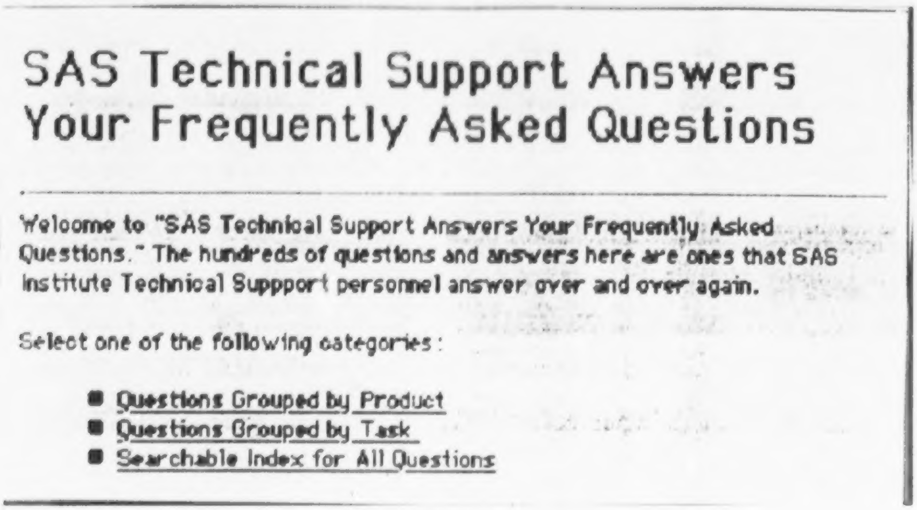
The Online Technical Support FAQ

The Technical Support section of the Customer Services home page has always been one of the most frequently accessed pages on the SAS WWW site. From this page, users can download support documentation, report a problem to Technical Support, or view libraries of online resources. Browsing the files in the Technical Support area is one of the quickest, most efficient ways a customer can find assistance.

From the Technical Support page, a user can bypass a phone conversation by opening a support request in EMITS (the e-mail interface to Technical Support). A user can browse over 10,000 SAS Technical Notes, downloading them as necessary; in contrast, the only other way of receiving this information is by fax or surface mail. Users that are not familiar with FTP can access the anonymous FTP site through their favorite web browser. This often saves time and aggravation on the part of the novice user.

If users are not sure what information they are looking for, they can use a searchable index to browse over 10,000 SAS notes. These notes reside in a large database and are indexed with keywords and titles. Likewise, they can access SIBBS (the SAS Institute Bulletin Board System) and share experiences with other users in the SAS community. Other resources on this comprehensive site include access to National Language Support information, an archive of the Technical Support mailing list, a SAS sample library, and access to the ever-popular Frequently Asked Questions (FAQ) list [Fig. 3].

BEST COPY AVAILABLE



SAS Technical Support Answers Your Frequently Asked Questions

Welcome to "SAS Technical Support Answers Your Frequently Asked Questions." The hundreds of questions and answers here are ones that SAS Institute Technical Suppport personnel answer over and over again.

Select one of the following categories:

- Questions Grouped by Product
- Questions Grouped by Task
- Searchable Index for All Questions

Figure 3: The Technical Support FAQ Page

The Technical Support FAQ was the first effort by SAS Institute to place a document online before placing the book in print. However, the popularity of the FAQ on the web was so impressive, it was quickly decided to keep this document only as an online version.

The format of a FAQ is ideal for web browsing. Hyperlinks from one topic can lead a user through a series of similar topics. The searchable index allows users to find key words in products they may not find otherwise. Most topics also have links to related documentation. With one click, the user finds themselves in the Publications Catalog, and able to place a quick order of a book. When necessary, there are links to our FTP site; users can download a file then move right back into the FAQ.

As the FAQ expands to over 1500 questions, we have found that this document has also become an invaluable resource to the employees of SAS Institute. By mirroring this document on our intranet, we have a tool that can be used by both SAS programmers and the Technical Support staff. The FAQ is updated as needed by members of our Technical Support staff.

The high usage of the FAQ stressed to the Technical Support staff that the WWW is no longer just a way to present printed information; it is a way to reach a larger audience by presenting web-specific facilities. In the future, the Technical Support Staff plans to implement an interactive question and answer section, a query for finding the status of a request, and a query to assign a Sales Representative when further purchases or site renewals are required.

Conclusion

While the SAS support resources available on the SAS WWW site are quite extensive, we continue to work with our offices around the globe to provide even greater specific information that is of relevance to the local SAS user. What began as a method to present "printed information" in an online format quickly evolved into a much larger project. SAS Institute's Web site has progressed to building web-specific facilities that will reach our worldwide audience.

SAS is a registered trademark of SAS Institute Inc.

BEST COPY AVAILABLE

VRML and the Web: A Basis for Multi-User Virtual Environments on the Internet

Wolfgang Broll
GMD - German National Research Center for Information Technology
Institute for Applied Information Technology (FIT)
D-53754 Sankt Augustin, Germany
email: wolfgang.broll@gmd.de
www: http://orgwis.gmd.de/~broll

Abstract: VRML is the basis for WWW based 3D data on the Internet. Although VRML was a static description language during its initial phase, it is currently extended to support interactive virtual environments and behavior of virtual world entities. In this paper we show, how this standard can be further extended in order to support multiple users and shared interactions. We will also show, how shared applications running on arbitrary servers may be integrated into this approach. Beside extensions of the language specification itself, extensions to browsers, servers and the communication protocol will be discussed. We will present a scalable solution based on multicasting.

1 Introduction

VRML -the Virtual Reality Modeling Language- is the standard for 3D data on the Web. In its initial version [Bell et al. 95] [Pesce 95] it was a static scene description language, without any possibilities to realize behaviors or user interactions. The scene descriptions (world files) are distributed over the Internet by HTTPD servers, using the standard HTTP protocol [Berners-Lee et al. 95], and displayed on the individual host by VRML browsers such as SGI's Webspace or our prototype SmallView [Plate 1]. But already after less than one year the VRML community voted on future extensions in the area of behaviors and multimedia extensions. For that reason six proposals [Broll et al. 96] [Mitra et al. 95] [VAG 96] for the new VRML 2.0 standard were submitted to the VAG (VRML Architecture Group).

Although there was a wide agreement upon the members of the VRML community to add multi-user support to future VRML specifications, most of the submitted proposals did not show, how this might be integrated into their behavior approach later on.

In this paper we will present the key concepts introduced in our proposal, going beyond interactions and behavior, in order to support multiple users and shared virtual environments based on VRML. We will focus on these additional extensions, which are rather independent of the final behavior specification and might be added in future releases of VRML. In the first part of this paper we will show, how the VRML specification can be extended to support avatars, shared applications as well as shared behavior and multi-user interactions. In the second part of this paper we will focus on the required network infrastructure, the distribution of avatars and virtual world entity updates.

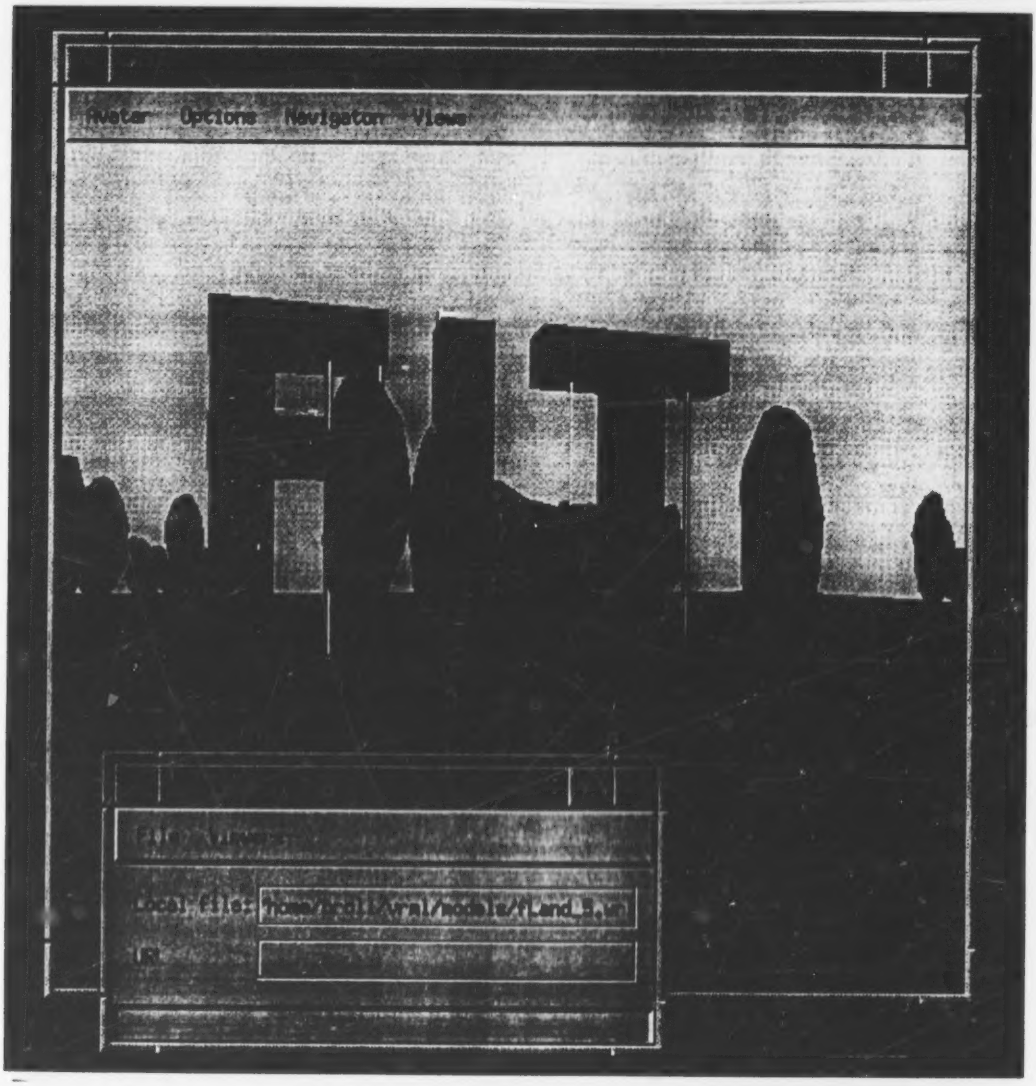


Plate 1: Screenshot of SmallView - our prototype of a multi-user capable VRML browser

2 Our VRML extensions

In this part of the paper we want to introduce some new nodes, we added to VRML in order to support multiple users and shared applications. The general structure of a VRML file is a tree of nodes. Each node consists of an identifier followed by a pair of brackets. Within these brackets node specific fields are specified. Grouping nodes allow to build up hierarchies of nodes by nesting them. Nodes are used to specify shapes (such as cubes, spheres, etc.), properties (such as material, position, orientation, etc.), etc. For a detailed discussion on the individual nodes types provided by the VRML specification please see [Bell et al. 95] [Ames et al. 96].

Avatars

To achieve a realistic impression of participating in a shared virtual world, other users as well as their current actions have to be represented in an adequate way. These user representations are called avatars or embodiments. Avatars may differ in shape and size. [Plate 2] shows three different avatars populating a single virtual world: two humanoid figures and an airplane.

BEST COPY AVAILABLE

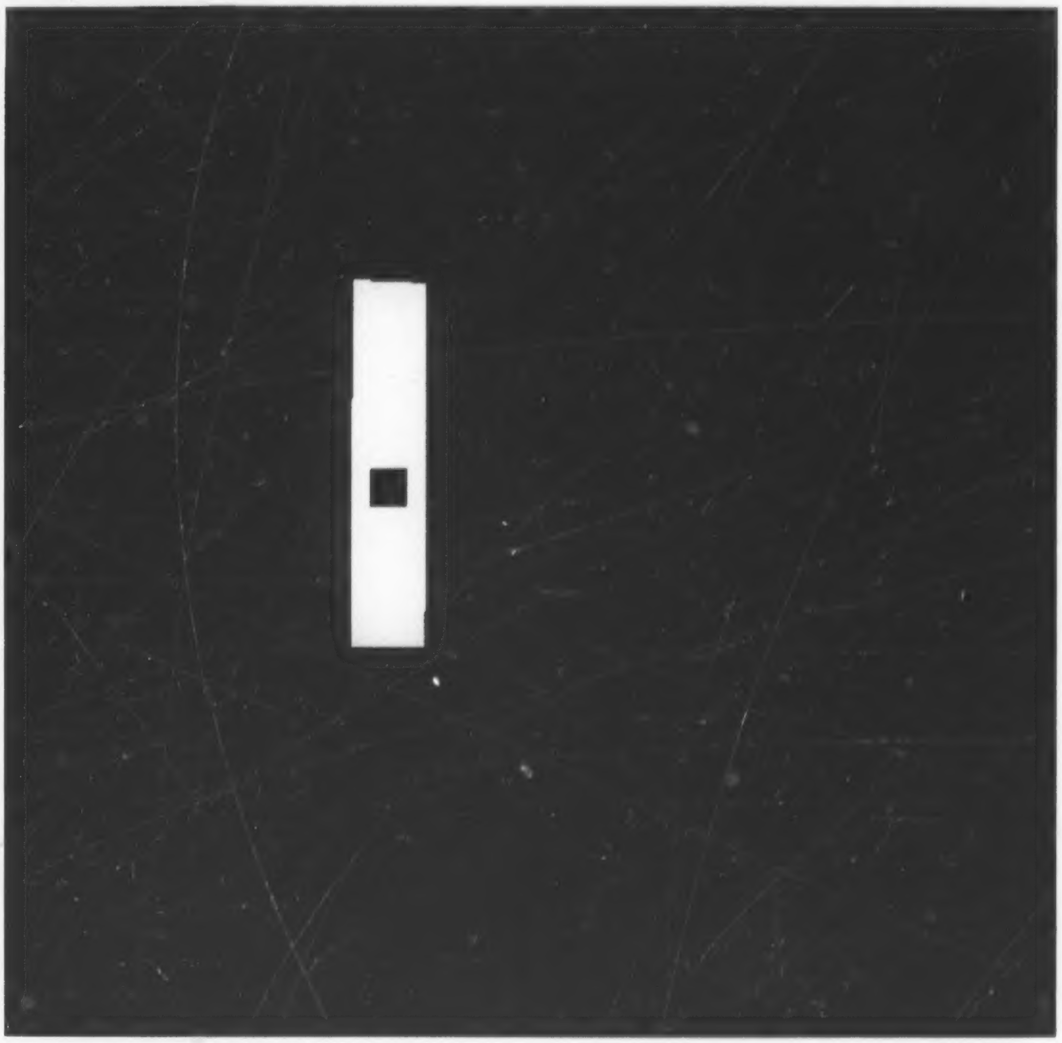


Plate 2: Three users represented by different avatars in a multi-user VRML world

Our extensions to VRML in order to support avatars, influence several aspects of user representations:

- the current avatar of the user
- alternative avatars of the user
- the user-specific and avatar-specific behavior
- groups the user belongs to
- navigation and camera settings for the user and/or a single avatar
- items which belong to the user or avatar, but are not visible to other participants (belongings)
- items the user or avatar is carrying

Each user needs a unique identifier, independent of the local virtual world. This is necessary in order to support user movements between different worlds (a user has to be identified as a certain user rather than just a user). A simple convention to achieve this, would be to use the user's email address or a user@host scheme as used for talk. Although users without avatars are not very useful in multi-user virtual worlds, we need to represent certain aspects of a user which are independent or at least might be independent of the currently selected avatar. Beside the user's id this includes the current location of the user (represented by a 3D transformation), his groups and avatar independent behaviors, items, and belongings. Groups can be used to define certain rights and/or restrictions for several users, independent

BEST COPY AVAILABLE

of their individual id. A user might participate in several groups.

[Fig. 1] shows the syntax of user and avatar nodes as realized in our current prototype. Within the brackets, the fields, their default values and the field type are listed. For a detailed information on fields see [Ames et al. 96].

Beside these avatar independent settings, each user needs one or several avatar nodes. These avatar nodes define the shape of the individual user embodiments, and optionally its relative location to the user's location, its type, camera and navigation style as well as avatar dependent belongings, items and behavior. Belongings are not part of the individual representations (shapes) and are neither visible nor accessible by other avatars (users) or the scene graph (unless we allow virtual robberies). In contrast to this, the item field is used to store objects the user may give to other users or drop in a world. This might e.g. be used to realize a kind of virtual shopping.

Default avatars may also be defined as part of the virtual world description. Further more the author of a virtual world might limit the possible user embodiments to a predefined set of avatars.

```
User {
  id ""                  # SFString
  groups ""              # MFString
  transform Transform    # SFNode
  belongings NULL        # SFNode
  items NULL             # SFNode
  <behaviors>            # depending on the final VRML spec.
  avatarType DEFAULT     # SFName
  avatarName             # SFName
  <avatars>              # different avatars of the user
}

Avatar {
  type ""                # SFString
  transform Transform    # SFNode
  camera NULL            # SFNode (PerspectiveCamera, # OrthographicCamera)
  navigationType NULL    # SFNode (NavigationInfo)
  belongings NULL        # SFNode
  items NULL             # SFNode
  <behavious>            # depending on the final VRML spec.
  <shape>
}
```

Figure 1: Proposed syntax of user and avatar nodes

Shared Applications

Shared applications are an important feature of distributed virtual environments. Beside the possibility to realize replicated applications, which are executed at each browser independently, our approach offers the additional possibility to run an application independent of the browser, potentially on a different server. While replicated behavior is transmitted as part of the virtual world description, either by specifying appropriate behaviors or by adding scripts to the scene (e.g. Java applets [van Hoff et al. 95]), external applications are completely independent of the virtual world. They communicate with entities of the virtual world by appropriate interfaces defined within the virtual world and as part of the browser.

In our approach, the part of the interface, which is defined by the virtual world is represented by a special Interface node.

BEST COPY AVAILABLE

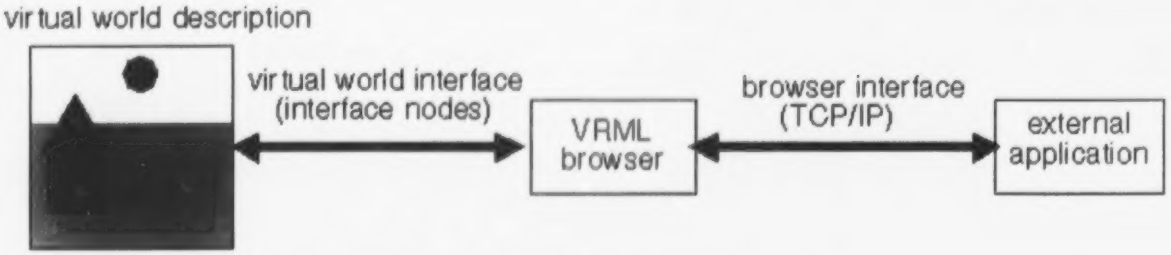


Figure 2: Interfae components to external applications

Interface nodes provide a simple but very powerful mechanism to establish connections between the VRML world and external applications. These applications might be located on the local host or even on arbitrary external servers. The communication is realized by a common event interface. Events sent to interface node are forwarded to the application or server by the VRML browser [Fig. 2]. Events sent to the browser by an application or server are sent to the Interface node. The communication between the browser and the external application is done by an TCP/IP interface of the browser. This interface is powerful enough for almost all kinds of applications. For external applications which have to communicate with a large number of hosts, a multicast connection (see second part of the paper) should rather be established. Locating an application on a central server which might be accessed from different VRML worlds allows to create high-level services. The Interface node allows behaviors defined within the scene graph to access such services.

The Interface node may be used for input, output or bidirectional I/O. The external service associated with the interface node, destinations of incoming events and the name of the external service can be specified by appropriate fields. The name of the server might be any valid host name or IP address and may include a specification of the port (e.g. baghira.universe.com:1234 or 134.62.47.11), providing the specified service.

A simple example for the usage of this mechanism is a 3D mail watcher connected to an external daemon by an Interface node. The daemon watches the user's mailbox and sends appropriate events to the interface. This will than raise the flag of the user's 3D mailbox within the virtual world.

Distributed Behavior

Our VRML behavior approach as well as any other proposal based on a user-extendable event mechanism, seems to be very suitable to support distributed virtual worlds. To do this, events influencing the state of a virtual world entity have to be transmitted over a network connecting the current participants of world. Our proposal [Broll et al. 96] uses Engine nodes to realize time dependent behaviors such as animations. The Moving Worlds proposal [Mitra et al. 95] uses TimeSensors instead. However, behaviors depending rather on the elapsed time than on user input allow more sophisticated ways to reduce the required network traffic [Roehl 95].

Some behaviors are even completely independent on each virtual world, e.g. a waterfall, a blinking light, tree movements by the wind, etc. These behavior do not need to be synchronized.

Other behavior have a defined starting and ending time, defined in real world time. No synchronization is needed for this kind of behavior either.

Further more most animations even if not completely independent are well defined. Thus their description can already be transmitted as part of the virtual world distribution. However, they may be invoked or stopped under certain conditions or parameters may be modified. This usually requires a single resynchronization. Since the transmission between the different sites requires a certain time (latency), such updates require a time stamp to guarantee the desired effects. Time stamps additionally require synchronized clocks, which can be realized using one of the well established mechanisms such as NTP (Network Time Protocol).

Our approach provides special engines nodes to allow the author of a virtual world to determine the required synchronization level. By an additional resynchronization field it is possible to determine, under which conditions a resynchronization is performed. Resynchronization may be deactivated, performed on activation or deactivation of the engine, or realized when any field value of the engine is changed. In the latter case, modifications are only distributed immediately if the engine is currently active, otherwise redistribution is postponed until the engine is reactivated.

Multi-User Interactions

In current proposals for VRML 2.0, interactions are executed by the local browser only, since multiple users or shared virtual worlds are not supported. However, within shared environments, inconsistencies may occur, when several users at different sites interact with the same virtual world entity. For that reason our approach adds shared trigger components to the VRML specification. These components either prevent several users to execute a certain behavior at the same time or detect and resolve concurrent access. Interactions might then be synchronized e.g. by combining them. Shared trigger components synchronize events, which would actually trigger a certain behavior. Synchronization is achieved by forwarding the events to the replicated copies of the trigger components. This allows one either to lock certain interactions already accessed by a another user, or to combine the input events of several users. We realize this kind of trigger components by two new nodes: BlockSharedTrigger nodes and MultiSharedTrigger nodes.

BlockSharedTrigger nodes as used to give a user an exclusive right to execute a certain behavior for a certain time. This mechanism provides a kind of lock which is automatically released, when the user deactivates the behavior or events from the locking site have not arrived for a certain time period. Only further events of the same sender (activator) lead to further executions of the behavior.

MultiSharedTrigger nodes are used to detect multi-user execution of a certain behavior. This node is used to trigger several events created by different users at their local browsers (located on different sites). It keeps track of all events with a time stamp within the specified time out interval. All received events are removed after the behavior has been executed. Nevertheless, this trigger component is be able to store further incoming events even before the execution of the behavior, since this execution might require more time than the specified time out value.

3 Our Server Extensions

In the second part of this paper we want introduce the necessary network and communication infrastructure in order to support large scale virtual environments on the Web.

In our prototype [Broll and England 95] we use a standard HTTPD server for the initial distribution of VRML files. However, these files are frequently updated by a second server daemon [Fig. 3]. This second server daemon (the multi-user daemon) is used for three purposes:

- adding and removing avatars of current users to the virtual world file
- updating, adding and removing virtual world entities of the VRML file
- supervising locks of virtual world entities and preventing clients (browsers) from unauthorized access

Since the original world file is modified by the multi-user daemon, even users of browsers which are not multi-user capable, can get a vague impression on the current state (including the current positions of avatars) of the virtual world. Multicasting Avatars of joining users, updates of virtual world entities as well as locking requests have to be distributed to the multi-user server daemon as well as to all clients. Realizing this by ordinary host-to-host network connections [Honda et al. 95], as used by HTTP, would exceed the capabilities and cause network overload. For that reason we use multicast groups to distribute the data. Multicasting [Kumar 95] provides a very flexible and powerful mechanism to distribute messages to a large number of hosts with minimal network load and distribution effort. Multicast packages are sent over a special part of the Internet -the Multicast backbone (MBone)- which connects multicast routers running a special multicast routing daemon (mrouted). Additional subnets can be connected to the MBone by tunnels (host to host connections). This structure allows the reduction of unnecessary message distribution on a very low level, which makes multicasting very effective for wide area communications of a large number of host, joining and leaving dynamically. Multicasting has already proven to be very suitable for this kind of application in existing distributed virtual environments [Carlsson and Christer 93] [Macedonia et al. 95].

In our approach a multicast group address and a port number identify a single world. This allows the daemon as well as the participating host, to receive messages from this group and port only. Future extensions might even use several groups or ports for update messages of a single world, in order to support the subdivision of large virtual worlds [Brutzman et al. 95]. The multicast address as well as the port number are transmitted to the clients (browsers) as part of the initial world distribution.

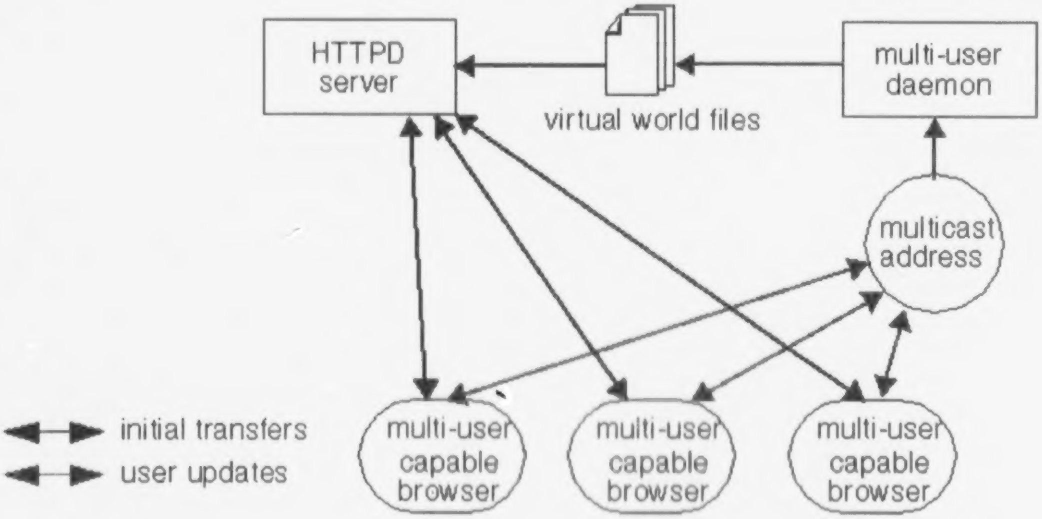


Figure 3: Avatar and update message distribution using multicast groups

Avatar Distribution

As already mentioned, avatars of users participating at a shared virtual world are added to the world file and distributed by the HTTPD server. However, usually each user will define his or her own individual avatar, or even several avatars of different types. These avatars are located at the host of the user (browser). When a user joins a virtual world, the appropriate avatar has to be transferred to all other participants as well as the multi-user daemon of the server. This is done via the multicast-group. If the user does not have an individual avatar or an avatar of the type required by the virtual world (this might be specified by the author), the multi-user server will distribute a default avatar via the multicast group.

Entity Updates and Locking

All updates on entities are sent to all other participants as well as the multi-user daemon via the multicast group address. Interactions on shared entities require a appropriate synchronization to guarantee the persistency of the virtual world. This can be achieved by appropriate behaviors to detect such interactions, as introduced in the first part of this paper, combined with an access control mechanism.

In our prototype this is realized by a very simple locking scheme, which requires very few network messages, since it does not require any acknowledge. Each participant can set a lock on an entity or behavior (shared trigger) by sending an appropriate message via the multicast group. Since multicasting is not order-preserving, concurrent locks may occur and lead to inconsistencies. To resolve these conflicts, the multi-user daemon acts as a watchdog, which determines the locking site. It then also sends a message to the multicast address, which allows all participating hosts to resolve the conflict and to restore persistency. Since such conflicts tend to be very sporadic, these additional messages are not significant for the total network traffic on the multicast group.

4 Conclusions and Future Work

In this paper we have introduced extensions to the emerging VRML standard in order to support large scale shared virtual environments. These virtual worlds are potentially populated by hundreds or even thousands of participants at the same time. By paying attention to the required communication infrastructure, we showed how VRML might be extended in the near future in order to support such virtual worlds. Our paper included solutions for individual user representations as well as for distributed behaviors and shared applications. In our future work, we will extend our current prototype implementation by additional features to support collaboration of multiple users in a more intuitive way.

5 References

[Ames et al. 96] Ames, L. A., & Nadeau, D. R., & Moreland, J. L. (1996). *The VRML sourcebook*. New York: John Wiley & Sons.

[Bell et al. 95] Bell G., & Parisi A., & Pesce M. (1995). The Virtual Reality Modeling Language, Version 1.0 Specification. See also http://vrml/wired.com/vrml.tech/

[Berners-Lee et al. 95] Berners-Lee, T., & Fielding R.T., & Nielsen H.F. (1995) Hypertext Transfer Protocol HTTP 1.0. HTTP Working Group. See also http://www.w3.org/hypertext/WWW/Protocols/Overview.html

[Broll and England 95] Broll W., & England D. (1995). *Bringing Worlds Together: Adding Multi-User Support to VRML, VRML'95 Symposium, 1995*, ACM SIGGRAPH, 87-94.

[Broll et al. 96] Broll, W., & England, D. , & Fechter, J., & Koop, T. (1996). The Power of Dynamic Worlds. Submitted VRML 2.0 Proposal. Also published as: *Towards Interactive Virtual Environments: Interaction and Behavior Extensions for VRML*. Technical Report WSI96-11, University of Tübingen, Germany. See also http://orgwis.gmd.de/VR

[Carlsson and Christer 93] Carlsson, C. & Hagsand, O. (1993). DIVE-A Platform for Multi-User Virtual Environments. *Computer & Graphics*, Vol.17, No. 6, 663-669. See also http://www.sics.se/dce/dive/dive.html

[Brutzman et al. 95] Brutzman, D. P., & Macedonia, M. R., & Zyda, M. J. (1995). *Internetwork Infrastructure Requirements for Virtual Environments, VRML'95 Symposium, 1995*, ACM SIGGRAPH, 95-104.

[van Hoff et al. 95] van Hoff, A., & Shaio, S., & Starbuck, O. (1995). *Hooked on Java,* Reading, Masachusetts: Addison Wessley. See also Java Language Documentation. Sunsoft (1995). See also http://java.sun.com/documentation.html

[Honda et al. 95] Honda Y., & Matsuda K., & Rekimoto J., & Lea R. *Virtual Society: Extending the WWW to support a Multi-user Interactive Shared 3D Environment, VRML'95 Symposium, 1995*, ACM SIGGRAPH, 109-116.

[Kumar 95] Kumar, Vinay (1995). *MBone - Interactive Multimedia on the Internet*, Indianapolis, Indiana: News Riders Publishing.

[Macedonia et al. 95] Macedonia M.R., & Zyda M.J., & Pratt D.R., et al. (1995). *Exploiting Reality with Multicast Groups: A Network Architecture for Large-Scale Virtual Environments, Virtual Reality*

Annual International Symposium, VRAIS `95, 1995, . Los Alamitos, CA: IEEE Computer Society Press, 2 -10.

[Mitra et al. 95] [Mitra, Yasuaki Honda, Kouichi Matsuda, Gavin Bell, Chee Yu, Chris Marrin: Moving worlds: behaviors for VRML. VRML'95 Symposium, (San Diego, Ca., Dec. 13-15, 1995) See also http://webspace.sgi.com/moving-worlds/

[Pesce 95] Pesce, Mark (1995. *VRML -Browsing & Building Cyberspace*, Indianapolis, Indiana: New Riders Publishing.

[Roehl 95] Roehl, Bernie. Distributed Virtual Reality - An Overview. June 1995. See also http://sunee.uwaterloo.ca/~broehl/distrib.html

[VAG 96] VAG - The VRML Architecture Group, http://vag.vrml.org

Tracking WWW Users: Experience from the Design of HyperVisVR

Chris Brown
Communications Research Group
University of Nottingham, University Park
Nottingham NG7 2RD ENGLAND
ccb@cs.nott.ac.uk

Steve Benford
Communications Research Group
University of Nottingham, University Park
Nottingham NG7 2RD ENGLAND
sdb@cs.nott.ac.uk

Abstract: We are generally concerned with the development of a hypertext database visualization framework, HyperVisVR, which supports3D collaborative visualization, dynamic responsiveness to database updates and representation of the underlying database users in the visualization. In developing a plug-in database access module for the World Wide Web (WWW), we have needed to solve several problems related to the WWW's loosely-coupled stateless architecture. One such problem is that of tracking WWW users as they move between pages and servers. This paper discusses current approaches to tracking WWW users, proposes new ones and explores issues of privacy and mutuality with respect to the monitoring of hypertext access.

Introduction

As the WWW increases in size and complexity, there becomes a need for advanced tools to manage information overload. Image maps, navigation bars, server push/pull, Java, ActiveX and Shockwave are among the layout techniques that are being used by content providers to generate the site designs which predominate the WWW today. Although these complex web site layouts may improve navigability within the site itself, when we consider the WWW as a global distributed hypertext database, they can cause confusion as browsing users are met with a bewildering array of non-standard navigational tools and controls. Despite this recent proliferation of these non-standard interfaces, hypertext navigability is not a new problem. In 1990, Jacob Nielsen addressed the issue of users becoming lost in hyperspace and proposed the use of overview diagrams (maps) and fish-eye techniques [Nielsen 1990]. These approaches have been adopted by several researchers trying to improve the navigability of the WWW, particularly Dmel with his WebMap tool [Domel 1994], and Mujherta with the Navigation View Builder [Mujherta 1994]. The motivation for this paper is a project entitled HyperVisVR, which is a framework for visualization of large hypermedia databases within a 3D virtual world. The driving goals of this framework are:

- Extensible through plug-in modules to support new databases and visualization styles.
- Fully dynamic visualization which is responsive to viewers movement through virtual world.
- Database access is represented in the visualization: users moving through the database have `embodiments'.
- Modifications to the database are reflected in the visualization.
- HyperVisVR applications support peer-peer communications to represent other visualization viewers in the visualization itself, and to share cached information about the database objects.

As an underlying database for HyperVisVR, the WWW has the benefit of being ubiquitously available throughout the world. However, we believe it poses some unique problems due to its loosely-coupled, stateless, architecture. This makes it a particularly interesting first target for our visualization framework. To be more specific, the problems with the WWW as a database for HyperVisVR are:

- It is difficult to extract rich meta information about objects on the WWW.
- Objects and links are stored together, making it difficult to analyse the structure of the WWW without performing an exhaustive search through the objects themselves.
- As a connectionless, stateless protocol, HTTP makes it very difficult to track users and their actions on the database on a global scale.

However, we don't believe that these problems are limited to HyperVisVR and many researchers are trying to find ways around the limitations on HTTP as a hypertext database protocol, and HTML as a hypertext object description language. Finding backwardly-compatible solutions to these problems would be of benefit to the research community as a whole. In this paper, we turn our attention to the specific problem of tracking users of the WWW as they browse and search the global WWW infrastructure. Potential applications of our ideas include:

- WWW providers might optimise their servers by analysing the paths of users through their servers.
- WWW providers might provide their clients with more detailed usage logs.
- WWW applications might provide a mutual awareness of users leading to opportunities for encounters.
- Tools to help users understand the relation between their current location and its surrounding context.
- Search tools might use the activity of users around a set of WWW resources to indicate interest in them.

User Tracking

We now focus on the problem of tracking users as they pass through the WWW. Users can be tracked at three different places. A server can track users moving through it, a browser can be modified to provide tracking information whenever a database access is made and proxy servers can track the users who use the proxy to access other areas of the WWW. The first two sub-sections below assume that all connections are direct between the user agent and the origin server. Proxies further complicate user tracking techniques, and are considered in section 2.3

Tracking at the Server

The key reason that user tracking at the server side is so difficult is that HTTP is stateless. Each request for an object from an HTTP server is completely separate, and cannot easily be associated with previous requests from the same user. The following sections discuss methods for user tracking that are available with current and proposed versions of HTTP, extensions to browsers that enable user tracking, and suggestions for improvements to HTTP which would make user tracking cleaner and more reliable.

In HTTP/1.0

There are several request headers specified in HTTP/1.0 [Berners-Lee et al. 1996] which can be used by the server to link up requests into a click-trail for a particular user in a given session. These are: From and Referer[sic]. In addition, the IP address of the machine generating the request can be used as a basis for tracking. The From request header is an optional field, specified to contain the Internet e-mail address of the user. However, only several less popular browsers actually make use of this field, as it is regarded by most to be a breach of the privacy of the user to send his personal information in each request. From fields are most widely used in WWW robots to identify the administrator in case there is a problem with the robot. The Referer request header can also help track click paths. This field can be used by the browser to specify the address of the resource from which the current request address was obtained. This can be interpreted by the server to link a previous request to the current one.

As browsers follow outside links from the server to other servers, their movement can be tracked through the use of cgi-redirect scripts. All outside links on a server are modified so that they link to a local cgi-redirect script, and pass the location of the outside link to the script as a parameter. The cgi-redirect script can log the movement, and then issue a redirect to automatically route the browser to the outside resource.

Some sites use intelligent algorithms which analyse log files and link up requests into click-trails through the server using a combination of IP address, the From and Referer request headers (if present), a maximum time between subsequent requests, and an understanding of server structure to recognise which items a user is likely to access based on his current position in the server. It may be possible to do this analysis in real-time as each request was served, but it would pose a large overhead to already over-loaded servers.

Authentication is often used by sites needing a fail-safe and universal way to track users through their servers. The advantages of this method are that it provides a method of user tracking which works with most browsers and it identifies individual users on subsequent visits even if they are connecting from a different computer. However, it is extremely inconvenient as it requires users to first register with the site and then remember a username and password that they should use on all subsequent visits to the site. Many users find that it generally isn't possible to use the same username and password at all such sites as many won't allow users their primary choice of username/password. Furthermore, based on our own experiences, we expect that many users who would otherwise look through the site may be dissuaded because of the inconvenience of registering or remembering their username. Thus, in general, we would not recommend this technique for sites hoping to get a high volume of traffic and who wouldn't otherwise use authentication for controlling access to their site.

HTTP/1.0 Extensions

The Cookie mechanism for client-side stateful transactions in HTTP is an extension to the HTTP protocol proposed by Netscape Corporation and implemented by the Netscape browser and several servers [Cookies 1995]. When a browser requests a resource from a server for the first time, the server responds with a cookie, which the browser stores and sends as part of each subsequent request. This allows the server to link up requests from a particular browser into a click-trail. Cookies can be persistent, linking requests from one browsing session with requests from the previous one.

The Keep-Alive extension to HTTP/1.0 allows several resources to be requested over a single connection. This is implemented by the Netscape browser and several servers. In principle it allows several requests to be matched up together as coming from one browser. However, in practice browsers use it only for single pages, requesting the page itself and all its embedded objects in one request. This limits the usefulness of the extension to follow a browser between distinct pages.

In HTTP/1.1

The HTTP/1.1 proposal introduces a new persistent connection architecture as the default connection type [Fielding et al. 1996]. This supersedes the Keep-Alive extension header described in 2.1.2. Any number of requests can be made on a single connection, until either the server or browser closes the connection. The specification does not make clear the circumstances under which connections should be closed or maintained and as, at the time of writing, there are no widespread implementations of the protocol, it is difficult to comment on whether this new architecture will improve user tracking. It is likely, however, that for matching of click-trails where the requests are punctuated by hours or days, that this architecture will not help.

Tracking at the Browser

The chief difficulty with server-side click-trail tracking using any of the mechanisms described in section 2.1 is that you can only track requests to the server. Frequently, browsers cache pages, and provide history mechanisms to allow navigation `Back' to the previous and `Forward' to the next page in the history. The browsers quite correctly do not generate new requests for this navigation. A consequence of this is, however, that it is not possible to maintain an accurate position of a user within a site if the user has navigated using the history mechanism. This is a major problem as a study at Georgia Institute of Technology analysed browsing strategies and determined that a total of 42.7% of navigation was through the history mechanism [Catledge et al. 1995].

An alternative to user-tracking at the server side is to extend users' browsers to send usage information to interested parties whenever a new page is accessed. This can be implemented very easily with Mosaic's Common Client Interface (CCI) and a small helper application which connects to the CCI port of the browser, and relays WWW movement information via TCP or multicast to interested parties. Both WebCast [Burns 1995] and FollowWWW [Brown et al. 1996] are applications which make use of this technique. An alternative would be a Netscape plug-in which monitors the actions of the user and sends movement information.

The major disadvantage of these schemes is that they need to be explicitly configured by the users. The Netscape plug-in needs to be downloaded and installed; the Mosaic CCI application needs to be downloaded, installed, and then Mosaic needs to be configured and the helper started at the start of each session. It is likely that due to this inconvenience few users will download and install the tracking software, unless some form of incentive is given (free access to a subscription service, etc.). Furthermore, there are then the problems shared by any public software developer: supporting multiple architectures, user support, fixing bugs and notifying users of updates.

Commentary

The loss of real-time server-side tracking accuracy due to history mechanisms described in Section 2.1 might suggest that a browser-oriented tracking mechanism would be more desirable. However, whilst browser-side tracking is certainly more accurate, it suffers from scalability problems, and potentially low uptake due to the explicit configuration required. It may be possible to increase the accuracy of server-side tracking with certain browsers through several techniques. One such technique is the use of an anchor to an unavailable background image, embedded in the page. Netscape and Mosaic's history mechanism causes all unloaded images to be re-requested whenever the page is revisited, so careful analysis of requests for this unavailable image, including the Referer header, will indicate when a particular page has been revisited through the history mechanism.

A hybrid approach which combines server-side and browser-side tracking would be for the server to include a reference to a small Java tracking applet with each page. The applet would have the sole responsibility of contacting the server each time the user departs a page or navigated back to a page through the history mechanism. The applet would be stored in the browsers cache, and so wouldn't be loaded across the network for each page. The Java applets may even be able to track the browsing activity levels on the workstation to determine if a user is actively viewing a page. Doing so, however, might be regarded as a breach of privacy.

Both of these approaches could be handled within the WWW server which issues the resources, or they could be served by `tracking servers', in a similar way to the current proliferation of `page counting servers'. Tracking servers could handle page tracking for a number of different sites. They could act standalone for their benefit only, or make the tracking information available to applications such as HyperVisVR through TCP, UDP or multicast communication.

Cache/Firewall problems

The need to accurately track the movement of browsers clashes horribly with the application of proxy cache servers and firewalls to drastically reduce the amount of network bandwidth consumed with redundant requests. The crux of the problem is based on the load-based algorithm proxy caches use to determine how long a particular object should be cached for. This results in the original WWW server `seeing' extremely few requests from proxy servers for its most popular resources (as the caches store these popular objects), and an disproportionately high volume of requests for its least popular resources. As the popularity of proxy caches increases, this could completely invalidate the use of visualization such as HyperVisVR, which relies heavily on usage and popularity information. Possible solutions to the proxy cache problem can be broadly categorised into `ignoring the cache', `beating the cache', and `working with the cache'. We now discuss each of these in turn.

It is possible to completely ignore the WWW population which accesses the server from behind a proxy cache by disregarding all requests which contain a `Via' header or the word `via' in the User-Agent header from the cache. Ignoring these misleading requests would seem like a good, straightforward approach to the problem. However, it is not possible to assume that the `direct access population' will be a representative sample of the entire requests. Many WWW users access through a cache because of organisation rules or country-based bandwidth problems, so by eliminating these users from the statistics you could unwittingly be excluding whole classes of users from the tracking statistics.

Many content providers have resorted to `beating the cache' when attempting to obtain full access statistics and tracking information. HTTP/1.0 specifies a `Pragma: no-cache' header, which is an instruction to the cache server not to cache that object. However, due to the misuse of the header by content providers, many proxy cache administrators have resorted to ignoring the header and caching the resource anyway. Cache busting methods include appending a random segment to the URL which confuses the cache into thinking that all the resources are different, and generating all pages dynamically through a cgi-script, the results of which do not get cached.

Working with the cache implies that you allow the proxy to cache your resources, and you are supplied with a log of cache-served resources originating from your server. This is not feasible without some form of automated system, as the number of proxy-cache servers is huge, and the structure of the proxies is completely uncontrolled. Such systems are under development, and when ready, it may be in the proxy cache's best interest to supply such information, as it might help the cache-busters in their defection from enemy to ally! To send information about each request to the origin server of every resource in the cache would severely impact on Internet bandwidth, and reduce the benefit of running a cache in the first place. It may be feasible however to consider some form of periodic batched transfer of information to selected origin servers that requested the information by setting a special header when the resource was first requested from the server.

It is worth taking a paragraph to consider the impact of proxy caches on the unavailable image and Java applet solutions presented above. Based on experiments with NCSA httpd 1.5.1, proxies do not cache 404 unavailable responses, which means that this solution will work to track users behind proxy caches and firewalls. Furthermore, it will not impact on the network as much as cache busting algorithms, because no data is being transferred - only the request for the unavailable image. Java applets are cached by the proxy, but they are still allowed to communicate directly with the tracking server. There have been some problems with running networked Java applets from within Netscape from behind firewalls. At the time of writing, the author knows no solution to this particular problem.

This concludes our discussion of current and possible techniques for tracking users as they access the World Wide Web. The following section now briefly discusses some of the ethical issues that arise as a result of this idea.

Privacy and Ethical Considerations

Techniques for tracking users, as discussed in this paper, raise a number of ethical issues concerning privacy and security. Such issues are complex and hardly ever clear cut. However, they are also of great importance and so warrant discussion and consideration during the technical development process. We begin by identifying the kinds of information that might be gathered about the presence and activity of users on the WWW. These include:

- Monitoring general access trends - recording patterns of access by groups of people.
- Anonymous monitoring of individuals - recording details of an individual's access but ignoring their identity.
- Non-anonymous monitoring of individuals - also recording the identities of people accessing the WWW.
- Persistent recording vs instantaneous awareness - deciding whether monitoring information is

recorded for subsequent storage, analysis and use or whether it is only made available at the time of access (e.g. to enable chance encounters and stimulate social interaction).

There are may also be many possible uses of such information. For example:

- Making colleagues generally aware of each other's presence in much the same way the shared buildings and offices support the coordination of activity through casual awareness between their occupants.
- Encouraging chance encounters between people browsing the same or related information.
- By information and service providers in order to enhance services (e.g. developing new paths through information based on analysis of patterns of use) or as part of billing and accounting.
- To enhance security by providing better awareness of who is accessing which sites and information (in the same way that video surveillance improves the security or many urban areas).
- It may be made available to third parties such as advertisers and government agencies.

It is important to point out that this kind of tracking may have clear positive benefits in some circumstances, but may have a negative impact in others. For example, information about an individual's preferences and activities might be used both to their benefit (to tailor systems to their needs) and also to their disadvantage (to compile personal profiles). Consequently, we propose three principles to guide the application and use of tracking techniques:

- Notification - visitors should be notified in advance of the mechanisms and policies that are in operation at a given site so that they can decide whether to visit or not.
- Mutuality - in general, systems should be designed so as to provide some degree of symmetry or mutuality as to awareness of presence. Thus, visitors to a site who are being monitored should be aware of the presence of their observers. Furthermore, it might be sensible to maintain a rough balance between the levels of information known by each party.
- Balance of power - both the observer and observed should be able to influence the level of awareness or monitoring. In particular, it might be argued that people should not be made visible without their consent and should not be able to become invisible without the consent of those around them.

Related Work to HyperVisVR

Related work fits into three main categories: visualizing hypertext databases, visualizing access to hypertext database servers, and mutual awareness mechanisms for users simultaneously accessing hypertext resources. Visualizing hypertext databases has been a hot topic at WWW and hypertext related conferences. The approaches can be grouped into two categories. The first is those that follow a user through the WWW and creating visual maps of browsing history, such as [Ayers et al. 1995], [Domel 1994] and [Takano et al. 1996]. Slightly different from these is HyperSpace which uses a 3D map to display browsing history, however it does not update the visualization automatically to show the current browsing position [Wood et al. 1995]. Another area of hypertext visualization research concerns the collection of a dataset from the hypertext database to visualize directly. 2D visualisations include venn-diagrams [Ralha et al. 1995], graphs [Mukherjea et al. 1995] and a distance-representing relief structure [Giradin 1996]. Virtual terrains (2.5D) have been used in the Hyper-G Harmony browser [Andrews 1995]. Full 3D visualisations include WebViz and GopherVR. WebViz uses a batch oriented approach to retrieve documents from a WWW server, parse them into a Hyperbolic tree structure, and display them [Munzner et al. 1995]. GopherVR is a 3D spatial interface to the Gopher system [McCahill et al. 1995]. Other research has focused on visualizing the access patterns on a particular server. Lamm et al. discuss a system which displays a globe in virtual reality, and maps the level of accesses to their WWW server from a particular geographical region onto the height of a bar extending from the appropriate location on the virtual globe [Lamm et al. 1996]. However caches, firewalls, and service providers upset their technique, as they mask the true origin of the browser.

Turning briefly to some of our own work in this area, the Internet Foyer is a virtual reality WWW visualization of a set of pages, with visual representations of browsing users moving over the structure as they access the pages and move between them. It supports mutual awareness of WWW users on the same or similar pages concurrently. It also has a link to the real world: a real-time image of the visualization is projected onto the wall of a real foyer, and a video wall within the visualization allows VR users to see back into the real foyer. As such it links three spaces: the real world, VR and the WWW, and can be though of as a "mixed reality" [Benford et al. 1996]. The Internet Foyer was the precursor of HyperVisVR, which adds fully dynamic visualization, peer-peer networking, generic hypertext database support and plugin visualization styles.

References

[Andrews 1995] Andrews, K. (1995). Visualizing Cyberspace: Information Visualization in the Harmony Internet Browser. First IEEE Symposium on Information Visualization, Atlanta, GA, Oct. 1995, pp. 97-104.

[Ayers et al. 1995] Ayers, E.Z., & Stasko J.T. (1995). Using Graphic History in Browsing the World Wide Web. Fourth International WWW Conference, 1995.

[Benford et al. 1996] Benford, S.D., Brown, C.C., Reynard, G.T., & Greenhalgh, C.M. (1996). Shared Spaces: Transportation, Artificiality and Spatiality. To appear in Proc. CSCW'96, Boston, November 1996, ACM Press.

[Berners-Lee et al. 1995]: Berners-Lee, T., Fielding, R., & Frystyk H. (1995). Hypertext Transfer Protocol 1.0. This is a work in progress, expiring on August 19, 1995. <draft-ietf-http-v10-spec-05.txt> (this is now RFC 1945)

[Brown et al. 1996] Brown, C., Benford, S., & Snowdon, D. (1996). Collaborative Visualization of Large Scale Hypermedia Databases, in CSCW and the Web, Proceedings of the 5th ERCIM/W4G Workshop. Arbeitspapiere der GMD 984, GMD, Sankt Augustin, April 1996.

http://orgwis.gmd.de/W4G/proceedings/visual.html

[Burns 1995] Burns, E. (1995). Collaborative Document Sharing via the MBONE.

http://www.ncsa.uiuc.edu/SDG/Software/XMosaic/CCI/webcast.html

[Catledge et al. 1995] Catledge, L.D., Pitkow, J.E. (1995). Characterizing Browsing Strategies in the WWW. Third International WWW Conference, 1995.

[Domel 1994] Domel P. (1994). Webmap - A Graphical Hypertext Navigation Tool. Second International WWW Conference, 1994.

http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/doemel/www-fall94.html

[Giradin 1996] Giradin L. (1996). Mapping the Virtual Geography for the World Wide Web. Poster Proceedings of the Fifth International WWW Conference, 1996.

http://heiwww.unige.ch/girardin/cgv/report/

[Fielding et al. 1996]: Fielding, R., Frystyk, H., Berners Lee, T. (1996). Hypertext Transfer Protocol 1.1. This is a work in progress, expiring on December 7, 1996. <draft-ietf-http-v11-spec-07.html>

[Lamm et al. 1996] Lamm, S.E., Reed, D.A., Scullin, W.H. (1996). Real-time geographic visualization of World Wide Web traffic. Computer Networks and ISDN Systems, 28(1996), 1457-1468

[McCahill et al. 1995] McCahill, M.P., Erickson, T. (1995). Design for a 3D Spatial User Interface for Internet Gopher. Proceedings of ED-MEDIA'95, pp 39-44, Graz, Austria, June 1995. AACE.

[Netscape 1995]: Netscape Corporation. Persistent Client State HTTP Cookies Preliminary Specification.

http://home.netscape.com/newsref/std/cookie_spec.html

[Mukherjea et al. 1995] Mukherjea, S., & Foley, J. (1995). Visualizing the WWW with the Navigational View Builder. Third International WWW Conference, 1995.

http://www.igd.fhg.de/www/www95/proceedings/papers/44/mukh/mukh.html

[Munzner et al. 1995] Munzner, T., & Burchard, P. (1995). Visualizing the Structure of the World Wide Web in 3D Hyperbolic Space.

http://www.geom.umn.edu/docs/research/webviz/

[Nielsen 1990] Nielsen, J.. The Art of Navigating Hypertext, Commuications of the ACM, 33, 3, pp. 297-310, ACM Press.

[Ralha et al. 1995] Ralha, C.G., & Cohn, A.G. (1995). Building maps of hyperspace. Internet Multimedia Information: WWW National Conference, 1995, Minho University, Portugal.

http://agora.leeds.ac.uk/spacenet/ghedini.html

[Spero 199?]: Spero, S. Progress on HTTP-NG.

http://www.w3.org/pub/WWW/Protocols/HTTP-NG/http-ng-status.html

[Takano et al. 1996] Takano, T., Kubo, N., Shimamura, H., Matsuura, H. (1990). A Directory Service Architecture for the World Wide Web. Poster Proceedings of the Fifth International WWW Conference, 1996.

[Wood et al. 1995] Wood, A.M., Drew, N.S., Beale, R., Heldley, R.J. (1995). HyperSpace: Browsing with Visualization. Poster Proceedings of the Third International WWW Conference, Darmstadt, Germany.

http://www.igd.fhg.de/www/www95/proceedings/posters/35/index.html

Published on the Web from the WebNet-96 (forthcoming) proceedings with permission of the Association for the Advancement of Computing in Education (AACE).

A Tool for Developing Adaptive Electronic Textbooks on WWW

Peter Brusilovsky
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
E-mail: plb@cs.cmu.edu

Elmar Schwarz and Gerhard Weber
Department of Psychology, University of Trier
D-54286, Trier, Germany
E-mail: {schwarz | weber }@cogpsy.uni-trier.de

Abstract: An Electronic textbook is a popular kind of educational applications on World Wide Web (WWW). We claim that adaptivity is especially important for WWW-based educational applications which are expected to be used by very different groups of users without assistance of a human teacher . In this paper we describe an approach for developing adaptive electronic textbooks and present an authoring tool based on this approach which simplifies the development of adaptive electronic textbooks on WWW.

1 Introduction

World Wide Web opens new ways of learning for many people. Now, educational programs and learning materials installed and supported in one place can be used by thousands of students from all over the world. However, most of the existing educational WWW applications use the simplest solutions and are much more limited than existing 'on-site' educational systems and tools. For many designers, an ideal format of educational WWW material seems to be a static electronic copy of a regular textbook: chapter by chapter, page by page, picture by picture. Such electronic textbooks are non-adaptive, i.e., students with different abilities, knowledge, and background get the same educational material in the same form.

We claim that adaptivity is especially important for educational programs on WWW which are expected to be used by very different classes of users without assistance of a real teacher (who usually can provide adaptivity in a normal classroom). Currently, we can name very few adaptive educational applications on WWW [Brusilovsky, Schwarz & Weber 1996; Kay & Kummerfeld 1994; Lai, Chen & Yuan 1995; Nakabayashi et al. 1995]. All these applications keep a model of the user between sessions and use it to adapt the teaching sequence and the presentation of the material to a given user. The problem is that adaptive electronic textbooks are "knowledge-rich" applications and they are not very easy to design. There are some authorinig tools for developing "static" electronic textbooks on WWW [Goldberg, Salari & Swoboda 1996; Thimbleby 1996], but there are no tools available to support a designer in creating an adaptive textbook on WWW.

A possible approach for designing adaptive electronic textbooks on WWW was suggested recently in [Brusilovsky 1995]. This approach was further elaborated by the ELM research group in the University of Trier which applied it for developing an adaptive WWW-based LISP textbook ELM-ART [Brusilovsky, Schwarz & Weber 1996][Schwarz, Brusilovsky & Weber 1996]. Now we have generalized the experience gained in ELM-ART project and developed a subject-independent tool which simplifies the process of creating adaptive electronic textbooks on WWW. In this paper we describe our current approach for developing adaptive electronic textbooks on WWW and present an authoring tool which is based on this approach. In addition, we provide some recommendations for possible users of our tool and a brief review of relevant works.

BEST COPY AVAILABLE

2 An Approach for Developing Adaptive Electronic Textbooks

Our approach to developing adaptive electronic textbooks on WWW based on the ideas from the areas of Intelligent Tutoring Systems [Wenger 1987] and Adaptive Hypermedia [Brusilovsky 1996]. Our adaptive textbooks use knowledge about its domain (represented in the form of domain model) and about its users (represented in the form of individual user models). The domain model serves as a basis for structuring the content of an adaptive ET. We distinguish two parts in an adaptive ET: a *glossary* and a *textbook*. Both these parts are based on the domain model. The student model is used by an adaptive ET to adapt its behavior to each particular user.

2.1 Content Structuring

The Domain Model and the User Model

According to our approach, the key to adaptivity in an adaptive ET are the domain model and the user model. The simplest form of domain model is just a set of domain concepts. By concepts we mean elementary pieces of knowledge for the given domain. Depending on the domain and the application area, concepts can represent bigger or smaller pieces of domain knowledge. A more advanced form of the domain model is a network with nodes corresponding to domain concepts and with links reflecting several kinds of relationships between concepts. This network represents the structure of the domain covered by a hypermedia system. The domain model provides a structure for representation of the user's knowledge of the subject. For each domain model concept, an individual user's knowledge model stores some value which is an estimation of the user knowledge level of this concept. This type of model (which is called an *overlay model*) is powerful and flexible: it can measure independently the user's knowledge of different topics.

The Glossary

The glossary is the central part of the ET. According to our approach, the glossary is considered as a visualized (and externalized) domain network. Each node of the domain network is represented by a node of the hyperspace, while the links between domain network nodes constitute main paths between hyperspace nodes. The structure of the glossary resembles the pedagogical structure of the domain knowledge and, vise versa, each glossary entry corresponds to one of the domain concepts. The links between domain model concepts constitute navigation paths between glossary entries. Thus, the structure of the manual resembles the pedagogic structure of the domain knowledge. In addition to providing a description of a concept, each glossary entry provides links to all book sections which introduce the concept. It means that the glossary integrates traditional features of an index and a glossary.

The Annotated Textbook

A human-written textbook represents human teaching expertise on how to introduce the domain concepts to the learners. It is usually a real textbook represented in hypermedia form. A textbook is hierarchically structured into units of different level: chapters, sections, and subsections. To make the textbook "more intelligent" and to connect it to the glossary, we have to let the system know what each unit of the textbook is about. It is done by indexing of textbook units with domain model concepts. For each unit, a list of concepts related with this unit is provided (we call this list *spectrum* of the unit). For each involved concept, the spectrum of the unit can represents also the role of the concept in the unit. Currently we support two roles: each concept can be either a *outcome* concept or a *prerequisite* concept. A concept is included in the spectrum as a outcome concept if some part of this page presents the piece of knowledge designated by the concept. A concept is included in the spectrum as a prerequisite concept if a user has to know this concept to understand the content of the page. Indexing is a relatively simple but powerful mechanism, because it provides the system with knowledge about the content of its pages:

the system knows which concepts are presented on each page and which concepts have to be learned before starting to learn each page. It opens the way for several adaptation techniques presented in the next subsection.

2.2 Functionality

Advanced Navigation

The knowledge about the domain and about the textbook content is used to serve a well-structured hyperspace. The system supports sequential and hierachical links between section. It generates the table of content where all entries are clickable links. In addition, it generates links between the glossary and the textbook. Links are provided from each textbook unit to corresponding glossary pages for each involved concept. On the other hand, from each glossary page describing a concept the system provides links to all textbook units which can be used to learn this concept. These links are not stored in an external format but generated on-the-fly by a special module which takes into account the user's current state of knowledge represented by the user's model. This approach is not only reducing page design time but also provides room for adaptation. In particular, our approach supports two adaptation techniques which have been applied in ELM-ART [Brusilovsky, Schwarz & Weber 1996]: adaptive navigation support and prerequisite-based help.

Adaptive Navigation Support

Our approach provides many more opportunities for browsing the course materials than traditional on-line textbooks. The negative side of it is that there is a higher risk for the user to get lost in this complex hyperspace. To support the user navigating through the course, the system uses *adaptive annotation*, an adaptive hypermedia [Brusilovsky 1996] technique. Adaptive annotation means that the system uses visual cues (icons, fonts, colors) to show the type and the educational state of each link. Using the user model, the system can distinguish several educational states for each page of material: the content of the page can be known to the user, ready to be learned, or not ready to be learned (the latter example means that some prerequisite knowledge is not yet learned). The icon and the font of each link presented to the user are computed dynamically from the individual user model. They always inform the user about the type and the educational state of the node behind the link. Red means not ready to be learned, green means ready and recommended, and white means no new information. A checkmark is added for already visited sections.

Prerequisite-based Help

The system knowledge about the course material comprises knowledge about what the prerequisite concepts are for any unit of the textbook. Often, when users have problems with understanding some explanation or example or solving a problem, the reason is that some prerequisite material is not understood well. In that case they can request prerequisite-based help (using a special button) and, as an answer to help request, the system generate a list of links to all sections which present some information about background concepts of the current section. This list is adaptively sorted according to the user's knowledge represented in the user model: more "helpful" sections are listed first. Here "helpful" means how informative the section is to learn about the background concepts. For example, the section which presents information about an unknown background concept is more informative then a section presenting information about a known concept. The section which presents information about two unknown background concepts is more informative then a section presenting information about one concept.

3 A Tool for Developing Adaptive Electronic Textbooks on WWW

3.1 Authoring with InterBook

Out tool (provisional name is InterBook) is aimed to help the author to transfer a normal textbook existing in electronic form into an adaptive ET. To get the most benefits from it an author should start creating an electronic textbook with a hierarchically structured MS Word file. This section demonstrates a typical scenario of using this tool for the case when the original textbook is available as MS Word file.

Step 1. Creating the list of domain concepts. Before starting to produce an adaptive ET, an author has to think about the list of domain concepts which will be used to annotate pages. An author does not have to have the list of *all* concepts before starting the work, it could be made when annotating the book.

Step 2. Structuring and Annotation. To let InterBook recognize the structure of a book, an author has to use the regular way of structuring an MS Word file. It means that the titles of the highest level sections should have a pre-defined text style "Header 1", the titles of its subsections should have a pre-defined text style "Header 2", and so forth. Then an MS Word file has to be annotated and indexed by the course author. An annotation is inserted into the file at the beginning of each section. The annotations have to be written using a special character style (hidden+shadowed). For each unit, the author can (but not have to) provide the set of outcome and background concepts. The format for the outcome annotation is: (out: concept-name1, concept-name2, ...).The format for the background annotation is: (pre: concept-name1, concept-name2, ...).

Step 3. Translation to HTML. The annotated MS Word file has to be saved in RTF formats and translated into an HTML file by the RTF2HTML program controlled by some specially designed settings. All annotations and section titles are translated into HTML comments which have a special format. The resulting "InterBook file" is just an HTML file annotated with several kinds of special comments.

Step 4. Parsing into LISP structure and Serving on WWW. When the InterBook server starts, it parses all InterBook files and builds the list of section frames. Each section frame contains the name and type of the unit, its spectrum, and its position in the original HTML file. The obtained LISP structure is used by InterBook to serve all the available textbook on WWW providing all advanced navigation and adaptation features. All content which the user will see on the screen is generated on-the-fly using the knowledge about the textbook, the user model, and HTML fragments extracted from the original HTML file. These features of InterBook are based on the functionality of he Common Lisp Hypermedia Server CL-HTTP.

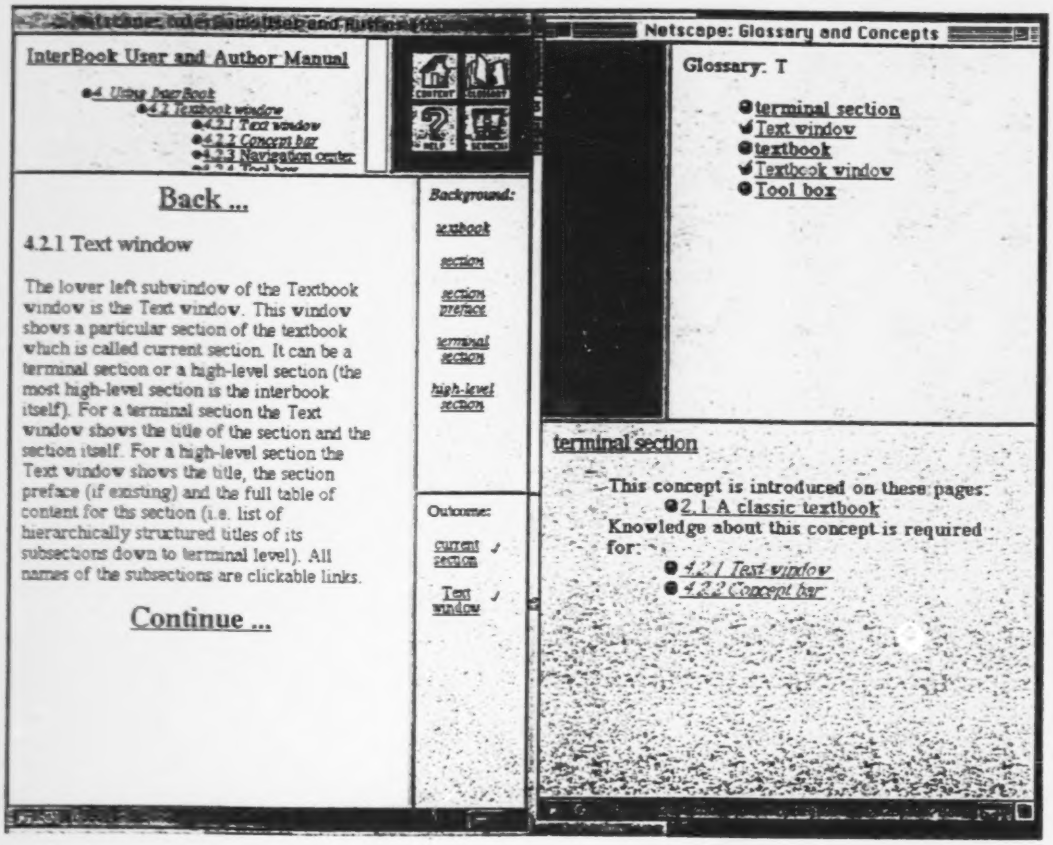
As we can see, our tool seriously simplifies the design of adaptive ET on WWW for the authors who use the approach presented in [2 An Approach for Developing Adaptive Electronic Textbooks]. It provides full support in preparation and serving an ET for the authors who know only how to use the MS Word text processor. An advanced used who have some knowledge on HTML and LISP programming can use our tool more flexibly. For example, an author can bypass step 1 and 2 by preparing the textbook directly in HTML format with annotations provided as specially formatted comments. The author can also replace server response functions and HTML generating functions to implement different structure and different "look and feel" of the be requested by a unique URL. To enable the server to respond to a particular URL, this URL has to be associated to a response function implemented in LISP which has to generate an HTML page on the fly as an adaptive response. CL-HTTP includes a set of LISP functions for generating pages.

3.2 Working with InterBook

InterBook is expected to be used with Netscape 2.0 or 3.0 browsers. It uses advanced features of these browsers such as multiple windows and frames to provide the user with useful and powerful interface. Main windows used by InterBook are the textbook window and the glossary window.

The Glossary window is used to view the glossary. The upper part of the window is a list of glossary concepts. The lower part of this window is used to show the glossary entry for a concept. For each concept the system presents the concept description (if provided by the author), the list of section titles (selected from all available textbooks) which present the concept (i.e., which have it as an outcome concept) and the list of section titles which require this concept (i.e., which have it as a background concept). Section titles are clickable links which makes the corresponding section to be loaded to the Textbook window.

The Textbook window is the most important window in InterBook interface. This window is designed to view the main content of a textbook, section by section. It is divided into frames performing different functions. Main frame of the Textbook window is the Text window. This window shows a particular section of the textbook which is called current section. For a terminal section the Text window shows the title of the section and the section itself. For a high-level section the Text window shows the title, the section preface (if existing) and the full table of content for the section (i.e. list of hierarchically structured titles of its subsections down to terminal level). A vertical bar to the right of the Text window is the Concept bar. It is used to show the concepts related with the current section. All names of concepts on the Concept bar are links to the Glossary. The upper part of the Textbook window hosts the navigation center and the toolbox. The navigation center shows the position of the current section in the textbook: it lists the titles of all direct predecessors (father, grandfather, etc.) and all brothers of the current section. All names of the sections are clickable links. The navigation center serves for both orientation and navigation. The toolbox provides a set of buttons which are used to call additional windows (such as content window, search window, and prerequisite-based help window) which provides additional functionality.



InterBook User and Author Manual

- *4 Using InterBook*
  - *4.2 Textbook window*
    - *4.2.1 Text window*
    - *4.2.2 Concept bar*
    - 4.2.3 Navigation center

4.2.1 Text window

The lower left subwindow of the Textbook window is the Text window. This window shows a particular section of the textbook which is called current section. It can be a terminal section or a high-level section (the most high-level section is the interbook itself). For a terminal section the Text window shows the title of the section and the section itself. For a high-level section the Text window shows the title, the section preface (if existing) and the full table of content for this section (i.e. list of hierarchically structured titles of its subsections down to terminal level). All names of the subsections are clickable links.

Background:

textbook

section

section preface

terminal section

high-level section

Outcome:

current section ✓

Text window ✓

Glossary: T

- ● terminal section
- ✓ Text window
- ● textbook
- ✓ Textbook window
- ● Tool box

terminal section

This concept is introduced on these pages:
- ● 2.1 A classic textbook

Knowledge about this concept is required for:
- ● *4.2.1 Text window*
- ● *4.2.2 Concept bar*

Figure 1: Textook and Glossary windows of InetrBook

4 Discussion: Indexing for "More Intelligent" Authoring on WWW

BEST COPY AVAILABLE

Authoring on WWW

The main idea behind our tool is using concept-based *indexing* to make conventional educational material more intelligent and flexible. The idea of indexing is to provide the information about the content of each unit of conventional educational material by indexing this unit with related domain concepts. Previously, indexing was applied in three authoring contexts: CAI context, hypermedia authoring context and ITS authoring context.

Indexing was originally suggested in CAI context by Osin [1976] who suggested a framework for indexing CAI frames by a set of topics which it covers. Such indexed sets of frames were not related to any pre-scribed order of presentation. They can be accumulated, stored in special libraries, and re-used by different authors to create their own courses. In the multimedia field, a similar idea of a re-usable database of multimedia learning material indexed by topics and keywords is elaborated by Olimpo et al. [1990].

Later indexing was applied in hypermedia and ITS authoring area. In the hypermedia authoring area, an idea of indexing was elaborated by Mayes, Kibby and Watson [1988] in the StrathTutor system. They stressed additional preference of indexing the frames of learning material - the possibility to indicate related pairs of frames not by tedious glossary linking of pieces of learning material together, but dynamically, on the basis of similarity of corresponded sets of topics. In the ITS authoring area, indexing was applied to turn traditional CAI into a "slightly intelligent" ICAI [Elsom-Cook & O'Malley 1990; Grandbastien & Gavignet 1994; Vassileva 1992]. "Slightly intelligent" ICAI are based on both the CAI and ITS paradigms. The teaching material is not generated as in 'orthodox' ITS, but stored in CAI-like frames. However, these frames are indexed with the concepts from an explicit domain model network, so they can be selected intelligently. The most recent application of indexing on the crossroads of the above directions is hypermedia-based ITS which use indexing technology to connect the learning material represented in a hypermedia form with the domain knowledge base: SHIVA [Zeiliger 1993], ITEM/PG and ISIS-Tutor [Brusilovsky, Pesin & Zyryanov 1993].

Indexing shows to be a relatively cheap and useful technology for authoring "more intelligent" hypermedia and CAI systems. We argue that it is the relevant technology for developing more adaptive and intelligent educational applications of WWW. Currently, we can name only ELM-ART [Brusilovsky, Schwarz & Weber 1996] as an example of a WWW adaptive electronic textbook based on indexing. We expect that the WWW will boost the research and development work on adaptive electronic textbooks. We hope that our tool will be useful for those authors who are interested to make their ET adaptive and to serve it on WWW.

5 References

[Brusilovsky 1995] Brusilovsky, P. (1995). Adaptive learning with WWW: The Moscow State University Project. In Held, P., & Kugemann, W.F. (Ed.), Telematics for Education and Training . Amsterdam: IOS. 252-255.

[Brusilovsky 1996] Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. User Modeling and User-Adapted Interaction, 6 (2).

[Brusilovsky, Pesin & Zyryanov 1993] Brusilovsky, P., Pesin, L., & Zyryanov, M. (1993). Towards an adaptive hypermedia component for an intelligent learning environment. In Bass, L.J., Gornostaev, J., & Unger, C. (Ed.), Human-Computer Interaction (Lecture Notes in Computer Science, Vol. 753). Berlin: Springer-Verlag. 348-358.

[Brusilovsky, Schwarz & Weber 1996] Brusilovsky, P., Schwarz, E., & Weber, G. (1996). ELM-ART: An intelligent tutoring system on World Wide Web. In Frasson, C., Gauthier, G., & Lesgold, A. (Ed.), Intelligent Tutoring Systems (Lecture Notes in Computer Science, Vol. 1086). Berlin: Springer Verlag. 261-269.

[Elsom-Cook & O'Malley 1990] Elsom-Cook, M.T., & O'Malley, C. (1990). ECAL: Bridging the gap between CAL and intelligent tutoring systems. Computers and Education, 15 (1), 69-81.

[Goldberg, Salari & Swoboda 1996] Goldberg, M.W., Salari, S., & Swoboda, P. (1996). World Wide Web - course tool: An environment for building WWW-based courses. Fifth International World Wide Web Conference.

[Grandbastien & Gavignet 1994] Grandbastien, M., & Gavignet, E. (1994). ESCA: An environment to

design and instantiate learning material. In de Jong, T., & Sarti, L. (Eds.), Design and production of multimedia and simulation-based learning material. Dordrecht: Kluwer Academic Publishers. 31-44.
[Kay & Kummerfeld 1994] Kay, J., & Kummerfeld, R.J. (1994). An Individualised Course for the C Programming Language. Second International WWW Conference "Mosaic and the Web".
[Lai, Chen & Yuan 1995] Lai, M.-C., Chen, B.-H., & Yuan, S.-M. (1995). Toward a new educational environment. 4th International World Wide Web Conference.
[Mayes, Kibby & Watson 1988] Mayes, J.T., Kibby, M.R., & Watson, H. (1988). StrathTutor: The development and evaluation of a learning-by-browsing on the Macintosh. Computers and Education, 12 (1), 221-229.
[Nakabayashi et al. 1995] Nakabayashi, K. et al. (1995). An intelligent tutoring system on World-Wide Web: Towards an integrated learning environment on a distributed hypermedia. ED-MEDIA'95 - World conference on educational multimedia and hypermedia, AACE. 488-493.
[Olimpo et al. 1990] Olimpo, G. et al. (1990). On the concept of database of multimedia learning material. World Conference on Computers and Education, North Holland. 431-436.
[Osin 1976] Osin, L. (1976). SMITH: How to produce CAI course without programming. International Journal on the Man-Machine Studies, 8, 207-241.
[Schwarz, Brusilovsky & Weber 1996] Schwarz, E., Brusilovsky, P., & Weber, G. (1996). World-wide intelligent textbooks. ED-TELECOM'96 - World Conference on Educational Telecommunications, AACE. 302-307.
[Thimbleby 1996] Thimbleby, H. (1996). Systematic web authoring. Symposium "The Missing Link: Hypermedia Usability Research & The Web".
[Vassileva 1992] Vassileva, J. (1992). Dynamic CAL-courseware generation within an ITS-shell architecture. 4th International Conference, ICCAL'92, Springer-Verlag. 581-591.
[Wenger 1987] Wenger, E. (1987). Artificial intelligence and tutoring systems. Computational approaches to the communication of knowledge. Los Altos: Morgan Kaufmann.
[Zeiliger 1993] Zeiliger, R. (1993). Adaptive testing: contribution of the SHIVA model. In Leclercq, D., & Bruno, J. (Eds.), Item banking: Interactive testing and self-assessment. Berlin: Springer-Verlag. 54-65.

Acknowledgments

Part of this work is supported by a Grant from "Alexander von Humboldt Foundation" to the first author and by a Grant from "Stiftung Rheinland-Pfalz für Innovation" to the third author.

A Comprehensive System to Develop Secure Web Accessible Databases

Robert Buccigrossi
Turner Consulting Group and the University of Pennsylvania, USA
butch@gradient.cis.upenn.edu

Albert Crowley
Turner Consulting Group, USA
crowley@www.tcg-inc.com

Daniel Turner
President - Turner Consulting Group, USA
tcg@us.net

Abstract: The World Wide Web (WWW) provides a unique opportunity for the development of database applications that are system independent and accessible to everyone on the Internet. Sometimes, though, an application is needed that has a distributed user base, but will contain data that should not be accessible outside a given set of users.

In this paper, we describe a system of developing applications with Web front-ends which constrain both the userbase and the allowable actions that individual users may take. This system is composed of three main concepts: the secure loop, which is a method of password-protecting Web pages, the swap, which retains the state of a user between execution of CGI scripts, and the lifetime of an input form, which is a modularization of the tasks that a general input form should complete. It is our hope that through the use of this system platform independent application interfaces will flourish on the Internet.

1. Introduction

The World Wide Web (WWW) [Berners-Lee et al. 1994] provides a unique opportunity for the development of applications that are system independent and accessible to everyone on the Internet. For example, if the Web were used as a front-end to a database, the capability of modifying and querying that database would become open to everyone with access to a Web browser. Sometimes, though, an application is needed that has a distributed user base, but will contain data that should not be accessible outside a given set of users.

In this paper, we describe a system of developing applications with Web front-ends that can be accessed only by a given set of users. A graphical interpretation of this system is presented in [Figure 1]. Between the user and the database lies the collection of Common Gateway Interface (CGI) scripts used to process form data, access the database, and create new pages to be displayed to the user. [NCSA 1996]

With CGI scripts, we develop several layers of routines to implement the database application. One layer, which we call the secure loop, is a method of verifying the user at every access. The swap is a method of retaining the state of partially completed forms. Finally, the query routines are the actual procedures that implement the database access.

This system has been used to successfully develop Web accessible databases for the National Institute of Health. [TCG 1995] Through this experience, we have created simple guidelines for designing CGI scripts for easier writing and maintenance. Finally, we compare this system of application development with other existing methods.

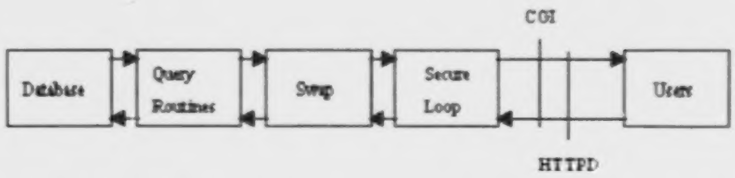


Figure 1: General System of Databases with Web Front-Ends

2. Review of Related Work

2.1 Web Accessible Database Authoring Tools

There are many tools available to create Web-accessible database applications. Many of these systems use some proprietary code combined with HTML. The most comprehensive authoring tools that we have seen at this time are Sapphire/Web [Bluestone 1995], Cold Fusion [Allaire 1996], WebBase [ExperTelligence 1995], and Web.SQL [Sybase 1996]. Each of these tools follow a slightly different path. Some focus on ease of database queries, or facilitate the transfer of form information to the scripts; others provide mechanisms to integrate HTML, SQL, and Perl. Still, none of these packages deal with every component of Web/database design.

It should also be noted that these products, when they support it, use proprietary techniques to maintain state and provide password protection. In this paper, it is our intention to explain and formalize some of the techniques that could have been used by these products to achieve their results. Future developers should see that it is possible to efficiently create complex databases without the use of external libraries or licensed code.

2.2 Maintenance of State and Password Protection

Unless information can somehow be associated with a user, Web documents are stateless, meaning that the execution of one CGI script will not effect the execution of another. [Gleeson & Westaway 1995] Two popular methods of imposing state are the use of hidden fields and the use of 'cookies.' [Netscape 1996a] The greatest danger with hidden fields is that a malicious user can view the HTML and modify the data in the hidden fields. Cookies stored locally can also be modified and are not supported by all browsers.

One solution is provided by a language called "Linda" created by Schoenfeldinger. [Schoenfeldinger, 1995] Linda records tuples among processes running in parallel on distributed systems. We simplify this concept by allowing CGI scripts to save tuples pertaining to state into what we call the "swap."

2.3 Password Protection of Document Trees

Many of the newer versions of Web servers allow password protection at a directory level, but this will not control how authorized users access the pages in that directory. MembersOnly is a system that allows document level access control. This system requires the use of a third party HTML server. [GATE Technology 1995] Neither of these systems permit conditional access to pages which may be required for database applications.

3. The Secure Loop

Every time a user tries to access a Web page in a secure database application, that user should be authenticated. The secure loop is a method to verify the user at every such access.

In the following discussion, we assume that the connection between the client and the server is secured using some protocol like Secure-HTTP (S-HTTP) [Rescoria 1996]. Discussion of the purpose and implementation of such security measures can be found in [Netscape 1996b].

3.1 Necessary Database Tables

The secure loop requires a mechanism to maintain the list of authorized users and a set of special values, called "tickets," used to associate a user with a specific set of screens they can access at a given time. Our implementation uses two relational database tables for this purpose: *AuthUsers* and *Ticket*.

The purpose of the table *AuthUsers* is to hold the list of authorized users, their passwords, and other information that could be useful to the application administrators. The purpose of the table *Ticket* is to hold the set of tickets that are valid for a specific user. It has at least three columns: *Username*, *TicketValue*, and *TimeStamp*. *Username* uniquely associates a set of ticket values with a user. The tickets held in *TicketValue* are random numbers generated at every Web access (see [Description of the Security Loop]). Ticket creation times are kept in the column *TimeStamp* so that they can become invalid after a specific amount of time.

3.2 Description of the Security Loop

[Fig. 2] displays the general interaction between the user, the Web server, and the scripts that process form data. The boxes in the shaded area represent actions that are done by all CGI scripts within the secure loop and can therefore be encapsulated in a library common to all of these scripts.

3.2.1 Initial Authorization

In our implementation, every user is given a password granting them access to the database. This password is encrypted and recorded into the *AuthUsers* table. When starting a database session, the first form the user encounters is a login screen. When the login form is completed, the encrypted password is compared with the *AuthUsers* table for that username. If the passwords match, the user is granted access to the application, a random ticket is generated, and the user enters the secure loop. Otherwise, an error message is displayed.

3.2.2 Inside the Secure Loop

Every CGI script in the secure loop uses the following procedure:

1. Check Ticket in Database: The script parses the command line extracting the username and ticket. It then executes a database query to ensure that the ticket is valid for the given username. If the ticket is valid, then the user is authorized to continue. Otherwise, an appropriate error message is given.

2. Process Form Data (if applicable): This is where the processing of the form data can be done safely. This could involve executing a query, updating tables in the database, or the like.

3. Generate Destination Ticket: A random ticket is generated, entered into the database and appended with the username to any outgoing link of the page created in step 4. Since the ticket is random and the link between client and server is secured, only the client and server know this ticket.

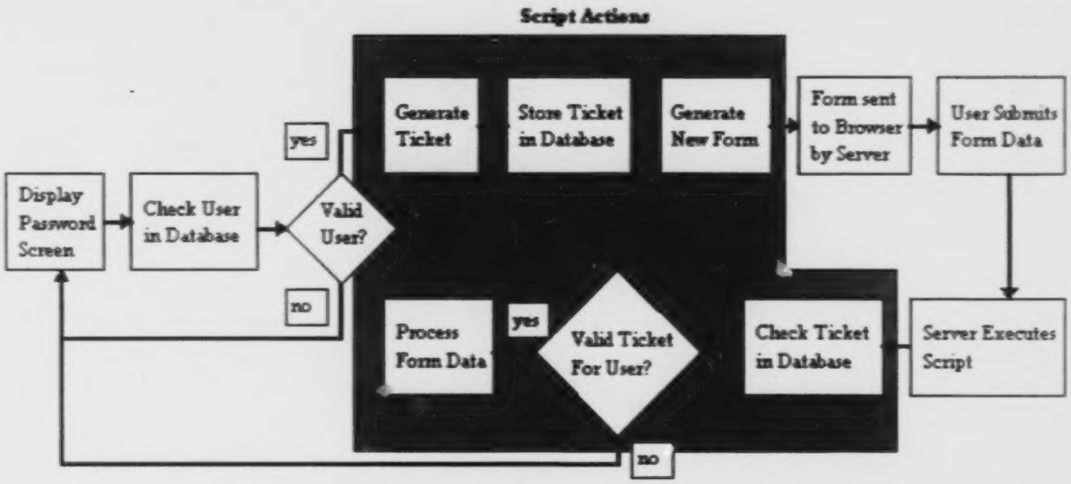

Script Actions
Generate Ticket
Store Ticket in Database
Generate New Form
Form sent to Browser by Server
User Submits Form Data
yes
Display Password Screen
Check User in Database
Valid User?
no
Process Form Data
yes
Valid Ticket For User?
Check Ticket in Database
Server Executes Script
no

Figure 2: The Security Loop

4. Generate New Form: Based on the results of the processed form or current state associated with the user, a new form is generated (for example displaying the results of a query or notifying the user of errors in a
form they have completed). The username and ticket must be added to any form destination or link definition so that the next script can execute securely.

5. User Interaction with the Form: The Web server passes the new form to the browser, which in turn displays it to the user. When the user submits the form, the form data is then passed back to the Web server which executes a new script to process it. This script begins the cycle again.

The only way that a user can access the application is by entering via the login screen. Because every other script must be accompanied by a username and a ticket there is no other way to enter the secure loop. Since we keep the usernames, passwords and tickets in the application database, the Web interface is as secure as directly accessing the database through other client/server interfaces.

3.2.3 Caching

Caching decreases the time required to access Web documents by temporarily storing them in an intermediate location. Pages copied to a local computer can be displayed more quickly than if retrieved from a remote site. If results of CGI scripts are cached by an Internet Service Provider or cached locally by a browser, sensitive information could be made available to anyone with access to the cache. Though recent developments in caching have reduced this problem [Gwertzan, 1996], timed invalidation of tickets can prevent an unauthorized user from entering the secure loop from the starting point of a cached document.

4. Form Consistency and Maintenance Issues

In a "user-friendly" application, invalid data entry should result in instructive error messages, or valid entry should result in context sensitive verification screens. Because of the inherent "statelessness" of Web applications, these common tasks are difficult to implement without somehow associating form contents and other important information with individual users. In this section, we introduce the "swap," which solves this problem by associating tag/value pairs of strings with a user and the current screen they are using. Also, we present methods to improve the ease of development and maintenance of scripts that implement input forms.

4.1 State and the Swap

BEST COPY AVAILABLE

State can be added to Web database applications through the addition of a table that we call the "swap" (since it is used to temporarily store form contents). Through use of the swap, tag/value pairs can be associated with a specific user and the screen they are currently accessing. The *Swap* table contains five columns: *Username*, *Screen*, *Type*, *Tag*, and *Value*. The *Username* must be a unique identifier, the *Screen* is a unique name given to each input form in the Web interface, the *Type* is used to specify different types of value pairs (for example form data or error messages), and the *Tag* and *Value* fields are used to record the state of the system (for example form field names and the data they contain).

When a user completes a form, the tag/value pairs sent to the CGI script are saved in the *Swap* table. If some form data form turns out to be incorrect, that specific region of the form can be reconstructed with appropriate help or error messages, leaving the remainder of the data stored in the swap. Also, if a user partially completes a form, but then presses a button to request help, by saving the form data to the swap, the form can be restored.

More importantly, the swap can be used to protect important data that should not be modified by the user. If a user modifies a row in a database table, it is important to retain the primary key of that row so that an update can properly be executed. If the primary key of the row is kept in a hidden field, the HTML code may be hacked so that the destination of the information is changed. By keeping the primary key in the swap, such a modification is impossible. Therefore the swap ensures the security of the application data.

4.2 Maintainability and the Lifetime of an Input Form

Code reuse can significantly reduce development time and cost. This can be through modularization of the tasks necessary to process an input form. The following steps are what we call the "lifetime" of a form:

1. Preparation - If the form is used to edit existing data, the necessary queries are made here. The results of the queries are saved into the swap to be used when the actual form is created.

2. Displaying the Form - Here the form is generated. If data for the form exists in the swap, then the data is used as the default values for the form objects. If error messages exist in the swap, these messages are displayed in appropriate places on the form.

3. Check the User's Input - If the user makes a request for a help screen, we display the help screen and then return to step 2. If there are errors in the data, we save error messages into the swap and then return to step 2. Otherwise we continue to the verification screen.

4. Verification Screen - If desired, the data that the user entered can be redisplayed for verification. If the user wishes to change something, we return to step 2. Otherwise we continue to step 5.

5. Update the Database - Here the data is removed from the Swap and entered into the Database. A screen is displayed that tells the user that the data has been committed to the database.

By identifying these elements common to many forms, routines can be written to encapsulate these steps. If separate CGI scripts are used for each step, scripts will not submit data to themselves. Our experience has found self-submitting scripts to be difficult to debug or modify and so they should be avoided at all costs.

5. New Developments

Recently, Java has changed the way many Web authors design sites and applications. In our Java applications, we use tickets and the swap to maintain high levels of security. Tickets are an excellent security mechanism for Java applets that do not retain a continuous socket connection with the database server. Also, the database server can use the swap to enforce proper behavior of the Java applets. A good programmer could decompile the Java byte code and rewrite an applet to modify internal state and allow illegal database access. By keeping the state in the swap, CGI scripts can ensure that the Java applets access only appropriate data items.

6. Conclusion

In this paper we have presented a system of developing applications with Web front-ends that can be accessed only by a given set of users. Using simple techniques like the secure loop, the swap, and the lifetime of an input form, the capabilities of commercial Web application builders can be duplicated and surpassed. These techniques have been found to be successful in the development of a number of secure databases for the National Institute for Health. It is our hope that by revealing these techniques, more users will be able to place their own applications on the Web and help stimulate the growth of the Internet to its full potential.

7. References

[Allaire 1996] Cold Fusion, Allaire. http://www.allaire.com/cfusion/

[Berners-Lee et al. 1994] T. Berners-Lee, R. Cailliau, A. Luotonen, H. Frystik Nielsen, and A. Secret (1994), The World-Wide Web. Comm. of the ACM, Vol. 37, No. 8, pp. 76-82, Aug. 1994.

[Bluestone 1995] World Wide Web Application Development, http://www.bluestone.com/papers/whitepapers/www_app_devel.html

[ExperTelligence 1995] ExperTelligence, Inc. WebBase - the Database Server of the Web, http://www.webbase.com

[GATE Technology 1995] GATE Technology, Member's Only - Security for the Web. http://www.iquest.net/cgi-bin/gate2?webmaster|gate/members.html

[Gleeson & Westaway 1995] Martin Gleeson and Tina Westaway (1995) Beyond Hypertext: Using the WWW for Interactive Applications. AusWeb95. Ballina, New South Wales, Australia. http://www.scu.edu.au/ausweb95/papers/hypertext/gleeson/

[Gwertzman 1996] Gwertzman, James and Seltzer, Margo. World-Wide Web Cache Consistency. http://www.eecs.harvard.edu/~vino/web/usenix.196/

[Netscape 1996a] Netscape, Persistent Client State Http Cookies. http://home.netscape.com/newsref/std/cookie_spec.html

[Netscape 1996b] Netscape, "On Internet Security" http://home.netscape.com/info/security-doc.html

[NCSA 1996] The Common Gateway Interface, http://hoohoo.ncsa.uiuc.edu/docs/cgi/overview.html

[Rescoria 1996] Rescoria, E. The Secure HyperText Transfer Protocol (Internet Draft) http://info.internet.isi.edu/in-drafts/files/draft-ietf-wts-shttp-02.txt

[Schoenfeldinger 1995] Werner J. Schoenfeldinger, WWW Meets Linda. Fourth International World Wide Web Conference. Boston, Massachusetts, USA. December 11-14, 1995. http://www.w3.org/pub/Conferences/WWW4/Papers/174/Overview.html]

[Sybase 1996] Sybase, Web.sql Programmer's Guide. Emeryville, CA: Sybase, Inc. http://www.sybase.com/products/internet/websql/docs/pguide/pguide1.htm

[TCG 1995] Turner Consulting Group, The Edison Database. http://era.info.nih.gov/

WEB ALGEBRA

Antonio Capani, Department of Computer Science,
University of Genova, Italy, capani@disi.unige.it
Gabriel De Dominicis, Department of Mathematics,
University of Genova, Italy, dedo@dima.unige.it

Abstract: In this paper we propose a model for a general interface between people and Computer Algebra Systems (CAS).The main features in our CAS interface are data navigation and the possibility of accessing powerful remote machines. Our model is based on the idea of Session Management. We implemented our model using the client-server capabilities offered by WWW. We envision our proposal as being useful for: educational purposes; computing; data navigation; communication of structured data; simulation of operations on distributed systems.

1 Introduction

Our goal is the construction of a common *World Wide Web* interface for different *Computer Algebra Systems* (CAS for short) particularly for systems devoted to *Commutative Algebra computations*. We need to use different systems because each of them is powerful for doing particular tasks. We can mention, among non-commercial systems developed by academic research groups, CoCoA[CNR], *GB*, *Macaulay*[MA], *Singular*[Sing] and the *PoSSo* server. All of them are capable of performing basic arithmetic operations on algebraic objects like: multivariate polynomials; rational functions; polynomial matrices; ideals. At the same time, advanced operations like Grobner bases[BU][RO] of ideals and modules, elimination, computation of syzygies and resolution of both ideals and modules are supported in all these systems.In addition some of these systems are designed to perform specific algorithms like, for instance, *Hilbert function*, *Hilbert driven Buchberger* Algorithm[GMRT] or *Tangent Cone Algorithm*[TC] in a particularly efficient way. These systems are mainly used as research tools but some of them are also currently being used for teaching purposes.

Because of these considerations it is clear that a common way to interact with the previously mentioned systems would be advantageous.

A prototypal version of our interface has been already demonstrated during the Workshop "Calcolo Simbolico" (November 20-21, 1995 Genova Italy) and during the Conference "Algorithmic Algebraic Geometry and Singularity Theory" (Schloss Dagstuhl, Germany, Jan 22 - 26 1996).

1.1 Interface requirements

The starting point of our investigation was our experience both as users and as (co)-workers on the CAS system CoCoA. In our experience we noted that the interaction between users and systems is characterized by mixture of hard computations (e.g. operations that involve Grobner basis computation) and data inspection. Data inspection (or better *data navigation*) is a key point in CAS interfaces since we usually work with huge and complex data. So we have to consider the possibility of different *views* of the same output.

Since many computations are quite hard it is sometimes necessary to use suitable computational resources that are not locally available. So we were challenged by the possibility of adding *client-server capabilities* to the interface.

Another good feature would be, of course, *independence of architectures* (machine, operating system, graphical interface) at least from the client side of the interface.

We focused our attention on the possible use of a WWW interface. This choice gives us an unconditioned independence from the architecture (a user simply runs some browser), hypertextual data navigation, automatic client-server capabilites (due to the use of HTTP servers) and some other advantages like a complete abstraction with respect to the computational resources we are interacting with.

To meet our requirements, we have organized the interactions between users and applications in sets of queries called *sessions*.The collection of all information about sessions and applications is called the Session Database (SDB for short).

The SDB is updated by a server, called the Session Manager SM for short), via a set of services: creation of new sessions; submission of queries; execution of applications. Starting from the SDB, the user can create reports to organize the outputs of the queries.

The SDB is stored on a single machine. Each operation is performed by running a process on the machine of the SDB.

Its important to note that our interface is interactive, that is, each query submitted to an application process may depend on the previous queries submitted to the same process. Moreover we can start a computation, turn off our computer, and then, possibly after several days, reconnect to the session to see the results.

The WWW interface to the SM is called WebSM. The WebSM is connected to the SM via a set of *cgi-applications*, one for each service. Furthermore, with the WebSM the reports can be hypertextual.

For use the hardest point to investigate was related to *client-server* capabilities of the interface. If we plan to build such a type of interface, first of all we have to solve some specific problems like *crash recovery*.

We will distinguish between local-crash recovery (usually due to net problems or browser crash) and remote-crash recovery (roughly speaking, crash of the SM or failure of cgi-applications).

2 The Session Manager

The Session Manager (SM for short) is the main engine of our tool. It is a server that enables interactions with the Session Database (SDB). WebSM is a WWW interface to the SM. Here is a bit of terminology:

An *application* is an executable program; an *application process* is an application in execution; an *application process agent* is a process on the same machine of the SM which passes the queries of the SM to the application processes; a *query* is a tuple `<i,o,Status,Infos>`, where `i` and `o` are respectively the input and the output of the query, `Status` is the current status of the query (see Sect. 2.3) and `Infos` contains auxiliary information (not specified here) used by the SM; a *session* is a set of queries; a *Sessions Database* (SDB) is a tuple `<A,P,S>`, where `A` is a set containing the information related to the applications connected to the SM (i.e. the machine in which the application runs, the agent of the application, etc.), `P` is a set of application processes and `S` is a set of sessions.

The user interacts with the application processes by means of the SM which offers the following services: `NewSession()` creates a new session; `AddQuery(s,i)` adds the query `<i,Null,Input,Infos>` to the session `s`; `StartAppl(s,a)` starts the application `a` within the session `s` by running a new application process; `SubmitQuery(s,q,p)` enqueues the query `q` of the session `s` to the queries for the process `p`; `RestartProcess(s,p)` starts a new process `p'` for the same application of `p` and passes all the queries queued for `p` to the new process `p'` (used in case of an abnormal termination of `p`); `RecoverProcess(s,p)` is the same as `RestartProcess()`, except that it also submits the previously answered queries to the new process in order to restore the state of the process as it was at the moment of the crash; `Report(Params)` creates a report according to the specified parameters (see Sect 2.4).

The SM solves the queries using the application processes. The SM takes a query `q` submitted by the user to a process `p` and sends the input to the agent of `p`. When a process has completed a computation, the SM puts the result in the output part of the associated query.

In the following sections we will examine, in more detail, how the SM and the WebSM work. We will support our presentation with a simple example, which we now introduce.We start with an empty SDB, and we consider two different sessions s_1 and s_2. The two sessions use two different processes, p_1 and p_2, for the same application (CoCoA), but they share the process p_3 for running the application *singular*. We will present the status of the SDB for the example, in a non-graphical style.

2.1 Sessions

The user submits requests to the SM within a session. Other users can share this session and exchange data within it; moreover different sessions can share the same process.

In the example below a new session has been created (service `NewSession()`).

```
==========================================================

Last operation: Created session 1 [29 Feb, 11:21:26]

------------------------------

Session 1 (no queries)

==========================================================
```

The user opens the connection with the WebSM by entering a session. The home page of the WebSM presents a list of available sessions, together with links that can be followed to enter each session.

2.2 Applications, Processes and Agents

An application is an executable program.The command `startAppl(s,a)` begins the execution of the application `a` using a new process `p`; it also stores in the Sessions Database (SDB) the information related to `p`.

The applications available via the SM can be on machines different from the one in which the SM operates.

The interaction between the process and the session manager is realized using *agents* that send the inputs to, and receive the output from, the application process via sockets, nfs or other forms of communication. The agents run on the same machine of the SM and communicate with it using the filesystem, in the following way: the SM writes the input of the query on a file (say fin), then the agent communicates the input to the (possibly remote) application, waits for the result and writes it on another file (say fout).

After the execution of the command `startAppl(1,cocoa)` the SDB looks as follows:

```
==========================================================

Last operation: Created process 1 [29 Feb, 11:26:09]

- Process 1 : (CoCoA) serving query 1.1 (history 1.1)

------------------------------

- Session 1 (1 queries)
```

```
1.1 Input: START CoCoA, Output: NULL, PID: 1, submitted

---------------------------------------------------------------
```

As the application process starts, a new query is automatically prepared to save the output related to the initialization of the application.

The WebSM presents to the user a page on which are listed all the applications available. For each application there is some information about the machine in which it resides (power and load of the machine, geographic location, etc.) and the list of the processes that run that application. Each running process has a link; clicking on it the user adds the process to the list of the active processes of the session. Moreover, the user can start a new application (command `startAppl()`) by activating the corresponding link ["START"]. The new process that runs the application will become, automatically, an active process of the session.

2.3 Queries

A query is a tuple `<i, o, Status, Infos>`.The user creates a query within a session `s` using the command `AddQuery(s,i)`, where `i` is the input of the query. A query can be submitted to a process using the command `SubmitQuery(s,q,p)` where `s` is the session, `q` is the query and `p` is the process.

Each query of a session has a status. Here is the list of all statuses and their meaning: **Input**: the query has an input, but it is not submitted to any process; **Submitted**: the input of the query has been submitted to a process, the identifier of the process and the submission time is contained in the `Infos`; **Running**: the process to which was the query submitted is now processing it; **Done**: the process has completed the computation and `o` contains its output;

```
---------------------------------------------------------------

Last operation: Created query 1.2 [29 Feb, 11:27:34]

- Process 1 : (CoCoA) serving query 1.1 (history 1.1)

------------------------------

- Session 1 (2 queries)

1.1 Input: START CoCoA, Output: NULL, PID: 1, submitted [960229,11:26:09]

1.2 Input: X := 1+1, Output: NULL, PID: NULL, input

---------------------------------------------------------------
```

The query 1.1 (the query 1 of the session 1) has been submitted to the process 1 which is now serving it. The query 1.2 is not yet submitted to any process.

Using the WebSM the creation of a query is done by writing down the input in a *text-area* and pushing the button ["NEW QUERY"']. The user can submit the query to a process by selecting it within the list of the active processes of the session.

```
---------------------------------------------------------------

Last operation: Served query 2.4 by process 3 [29 Feb, 12:01:39]

Process 1 : (CoCoA) ready (history 1.1 1.2 1.4)

Process 2 : (CoCoA) ready (history 2.1 2.2 2.3)
```

BEST COPY AVAILABLE

```
Process 3 : (Singular) ready (history 1.3 1.5 2.4)

-------------------------------

Session 1 (5 queries)

1.1 Input: START CoCoA, Output: Welcome to CoCoA, PID: 1, done

1.2 Input: X := 1+1, Output: OK, PID: 1, done

1.3 Input: START Singular, Output: Wilkommen zu Singular, PID: 3, done

1.4 Input: X, Output: 2, PID: 1, done

1.5 Input: X := x+y, Output: OK, PID: 3, done

-------------------------------

Session 2 (4 queries)

2.1 Input: START CoCoA, Output: Welcome to CoCoA, PID: 2, done

2.2 Input: X := 1+2, Output: OK, PID: 2, done

2.3 Input: X, Output: 3, PID: 2, done

2.4 Input: X, Output: x+y, PID: 3, done

==============================================================
```

In the query 1.2, the user asked the process 1 to compute the expression 1+1 and to store the result in the variable X.

In the query 1.4 the user asked the process 1 to print the value of the variable X.

Something similar happened for the queries 2.2 and 2.3. The outputs of the two (identical) queries 1.4 and 2.3 are different because the queries have been submitted to two different processes, even if they run the same application.

The interaction with p_3 shows how two users can communicate data by sharing the same process. The first user puts in the variable X of p_3 the polynomial x+y (query 1.5) and the other user looks at it by asking the process p_3 for the value of the variable X (query 2.4).

2.4 Reports

An important service of the SM is the production of reports. Each user can ask the Session Manager for a report on one or more sessions. Different users can ask for different reports on the same SDB and on the same sessions.

The reports may differ from the order of presentation of the queries(by submission time or by creation time), by the portion of the view (query of a process, query of all the processes of an application, query range). Furthermore the user can ask for a report containing just the output of a particular query. Often the outputs are very big and the user would prefer to navigate them in a hypertextual style. For example:

a matrix of polynomials could be reported as a matrix (HTML table) of hyperlinks, each of which can be represented by a string that summarizes the polynomial in that entry (e.g.: it gives the length of the polynomial). The user interested in seeing the entire polynomial can click on the link and receive the document containing the polynomial.

As WebSM runs on a WWW client, the navigation on the data is easy to perform. Obviously the applications should give sufficient information to the WebSM for the production of hypertextual outputs (for example in HTML). It is interesting to note that the applications can cooperate with their agents in producing hypertextual documents. The HTML makes it easy to write documents in "pretty" style. For example the HTML 3 will support "mathematical environment".

2.5 Scheduling of the Queries

The SM communicates each submitted query to the associated process by running the agent of the process. Periodically the SM checks to see if some process has completed a query (i.e.: its agent has written the result on the output file); if so, the SM copies the result of the query in the field `output` of the query itself, and sets the status of the query to **Done**.

Then, the SM gets the next query queued for that process and contacts the agent for its computation.

2.6 Crash Recovery

Since all the sessions are stored in the machine on which the SM runs, the crash of a client's machine is not a problem. The problem of crash and its recovery is only a problem on the SM machine and the machines which run the processes.

The use of different applications on different machines implies a higher probability of a crash. The SM model provides a simple way to do crash recovery of abnormally terminated processes. The list of all inputs computed by each process (i.e. the history of the process) is stored in the part of SDB devoted to the processes;

when a process abnormally terminates its execution, we can restore it by submitting all the queries processed until its termination.

E.g.: suppose that after the computation of the query 1.2 the process 1 dies.

```
------------------------------------------------------------

Last operation: Process 1 abormally terminated [29 Feb, 14:17:42]

Process 1 : (CoCoA) defunct (history 1.1 1.2)

------------------------------

Session 1 (4 queries)

1.1 Input: START CoCoA, Output: Welcome to CoCoA, PID: 1, done

1.2 Input: X := 1+1, Output: OK, PID: 1, done

1.3 Input: X, Output: NULL, PID: 1, submitted

------------------------------------------------------------
```

We can restart the process by submitting the command `RestartProcess(1,1)`.

```
------------------------------------------------------------

Last operation: Served query 1.3 by process 2 [29 Feb, 14:17:59]

Process 1 : (CoCoA) defunct ... 1.1 1.2
```

BEST COPY AVAILABLE

```
Process 2 : (CoCoA) ready ... 1.4 1.3

------------------------------

Session 1 (5 queries)

1.1 Input: START CoCoA, Output: Welcome to CoCoA, PID: 1, done

1.2 Input: X := 1+1, Output: OK, PID: 1, done

1.4 Input: START CoCoA, Output: Welcome to CoCoA, PID: 2, done

1.3 Input: X, Output: Undefined var X, PID: 2, done

=============================================================
```

We note that the query 1.4 has been moved to the new process 2; but that the process 2 has no memory of the assignment done in the query 1.2. We can perform crash recovery by using the command `RecoverProcess` instead of using the command `RestartProcess`.

```
=============================================================

Last operation: Served query 1.3 by process 2 [29 Feb, 14:18:35]

Process 1 : (CoCoA) defunct (history 1.1 1.2)

Process 2 : (CoCoA) ready (history 1.4 1.2 1.3)

------------------------------

Session 1 (5 queries)

1.1 Input: START CoCoA, Output: Welcome to CoCoA, PID: 1, done

1.2 Input: X := 1+1, Output: OK, PID: 2, done

1.4 Input: START CoCoA, Output: Welcome to CoCoA, PID: 2, done

1.3 Input: X, Output: 2, PID: 2, done

=============================================================
```

Note that the process has now recomputed the query 1.2 and hence the query 1.3 is correctly answered.

Obviously this approach is not efficient but it is fundamental because it is transparent to the user. Remember that our goal was to produce a common, user-friendly interface to different applications.

Moreover, depending on the application used, the user can speed up possible crash recoveries by "saving the state" of the process and possibly "restoring it'". Suppose for example that we do some big computation and store its result in a variable, using a command like `x := Hard(A,B,C)`. Also suppose that `v` is the final value of the execution. Then a simple way of helping the SM is to replace the query `x := Hard(A,B,C)` with the new one `x:=v`. In this way if a crash occurs the hard computation will not be redone.

2.7 Educational issues

Since many users can share the same session, our model is useful for teaching purposes. E.g. we can suppose that one user (the teacher) writes down a session and that other users (the students) read it. In this way the WebSM can be used for giving a lesson. Suppose now that the teacher proposes an exercise to the students. The teacher can then write down one query for each student and ask each student to write

BEST COPY AVAILABLE

down the result of ther investigation. Then the session will contain all the exercises written by all the students. We are certain that educators will find many other interesting ways to use the potentiality of a system constructed in this way.

The shared use of a session implies that considerations of ownership of sessions (and of queries inside a session) and of security must be made. For example one might prefer that the only person who can write the lesson is the teacher. Furthermore we want to avoid the possibility that a student could look at the exercise done by another student. A simple solution to this problem is to add the possibility for users to declare themselves, with a password and thus protect their data. This aspect of the SM (and of the WebSM) is currently under development.

REFERENCES

[BI] A.M. Bigatti (1996). Computations of Hilbert-Poincare' Series, To Appear on Journal of Pure and Applied Algebra.

[BU] B.Buchberger (1985).Grobner bases: An algorithmic method in polynomial ideal theory. Recent Trends in Mathematical Systems Theory, (N.K. Bose, Ed.), D.Reidel, Dordrecht, 184-232.

[CNR]A.Capani, G.Niesi, L.Robbiano. (1995) CoCoA, a system for doing Computations in Commutative Algebra,.Available via anonymous ftp from *lancelot.dima.unige.it*

[MA]Bayer D. and M. Stillman (1982-1990). Macaulay: A system for computation in algebraic geometry and commutative algebra. Available via anonymous ftp from *zariski.harvard.edu*

[TC] Pfister, G. The tangent cone algorithm and some applications to local algebraic geometry. In Mora, T. and Traverso, C. (eds.),Effective Methods in Algebraic Geometry, Progress in Mathematics 94, Birkauser Verlag, Basel, 401-409.

[RO] L Robbiano (1988). Introduction to the theory of Grobner bases, Queen's Papers in Pure and Applied Mathematics.

[GMRT] P.Gianni, T.Mora, L.Robbiano, C.Traverso (1994). Hilbert function and Buchberger algorithm preprint.

[Sing]G.M. Greuel, G.Pfister, H.Schonemann, Singular, A Computer Algebra System for Algebraic Geometry and Singularity Theory.

Acknowledgements

Thanks to Gianfranco Niesi and Lorenzo Robbiano for their support during the preparation of the manuscript. Thanks to Tony Geramita for (grammatical) help with the final draft.

Website Analysis: A Scaleable, High-Performance Software Platform for Powerful and Flexible Activity Recording, Accessing, and Reporting

July 1996

Scott Capdevielle
Chief Technology Officer
Andromedia, Inc.
545 Mission Street
2nd Floor
San Francisco, CA 94105
scott@andromedia.com
Tel: 415-278-0708
Fax: 415-278-0719

Table of Contents

Executive Summary 3Objective 3Chapter One: The Web Marketplace 4Growth Projections 4The Market 6Chapter Two: System Requirements and Specifications 9Technical Requirements 9Chapter Three: Andromedia's Aria Web Recording,Accessing and Reporting Platform 12Introduction 12Primary Design Choices 12The Aria Recording and Reporting System 15Administration Tools 24Developer's Toolkit 24Summary/Conclusion 25Table of Technical Requirements vs System Features 25

Executive Summary

The Internet is experiencing unparalleled growth. Managing the growth effectively requires a truly scaleable, high performance interactive activity recording, accessing and reporting system.

Current solutions are not able to satisfy market requirements. These solutions lack scaleability, flexibility and real time information.

Andromedia's Mission Statement is to develop and market the most powerful, flexible and widely used software platform for recording, accessing and reporting anonymous activity on the Internet, intranets, and related media.

Objective

This white paper details the underlying architecture of the Andromedia platform and explains the choices made in building this system.

Chapter One: The Web Marketplace

Growth of the interactive world as an electronic infrastructure for business, commerce, marketing, education and entertainment is unparalleled.

Growth Projections

The astronomical growth of the World Wide Web (Web) has been well chronicled. Accepting the analysis of many market research companies (such as the Gartner Group, Hambrecht and Quist, Morgan Stanley, Goldman Sachs, Input and many others) as having concluded that the Web is *already* a significant factor in the global electronic economy, Andromedia believes the most significant growth in use and dependence on the Web is just in front of us. We believe the activity level seen on the top ten Websites of *today* will be only the *average* in less than 18 months. In less than 18 months, companies and individuals will be more dependent on the Web for conducting their professional and personal lives efficiently and productively.

We conclude that high website growth is due to:

1) Content

2) Bandwidth

3) Access

4) Advertising

1) Content Authoring Tools

Content authoring tools are becoming much more powerful, easy to use and abundant.

Recently, the industry has seen a barrage of powerful and easy to use tools and languages arrive on the market, dramatically changing the content creation business. These new tools range from entirely new languages for developing content (such as Java from JavaSoft, a Sun Microsystems Company) to added features in existing tools such as the "save as" function in Microsoft Word. Shockwave from Macromedia and other multimedia tools have also contributed to this revolution.

The result of this step-function improvement in capabilities of available tools is leading to the creation of more interesting and useful content.

2) Increased Communication Bandwidth

Corporate users generally have a better experience than home users, as they can access the Web through higher speed corporate networks and dedicated T1 or T3 data lines. The home user is still limited to the capability of the twisted-pair copper phone line and the 19.2kbps modem which is currently available.

The fact that most Web users today spend on average 80% of their on-line time *idle* makes for a very inefficient use of the Web.

There is a tremendous amount of activity in the telecommunications arena, all aimed at essentially making bandwidth virtually unlimited. Cable companies are upgrading their plant infrastructure to deliver digital information over the coaxial cable which today is delivering only analog TV broadcast. Modem manufacturers are gearing up on producing cable modems. The RBOCs (regional Bell operating companies) are all offering new services based on more efficient modulation schemes delivered over existing copper wiring. These include ISDN and ADSL, which will improve communication speeds by up to 30 times the 19.2kbps speeds supported currently.

The buildout of the long haul network (SONET) backbones based on ATM, running over fiber optic communication cables is underway. As the reach of fiber optic cable gets closer to the home and the workplace (FTTH, FTTC), speeds of communicating data will increase by more than 2 orders of magnitude.

Users will reach optimal efficiency, at which time they will spend little of the current 80% of their time waiting on the network to serve up the information they seek. This element will contribute a factor of 4x or 5x to the increase in Website activity levels.

3) Ubiquitous Access to the Web

A fully web-capable PC now costs about $1,000, including all of the software needed to connect to the Web. Companies such as AT&T have announced virtually free connection access to the Web, which previously cost $200 to $300 per year. In current dollar terms, PCs are less expensive than color TVs were at that point in time when color TVs became ubiquitous appliances in the home. In addition, the process of getting connected to the Web has been greatly simplified with the emergence and acceptance of plug-and-play standards, as well as tightly bundled packages that offer the full set of services needed to connect to the Web as a user or a Website. ISPs (Web Service Providers) have proliferated and will provide full turnkey system connection via a single point of contact. These advancements have significantly dropped the barrier of getting connected to the web.

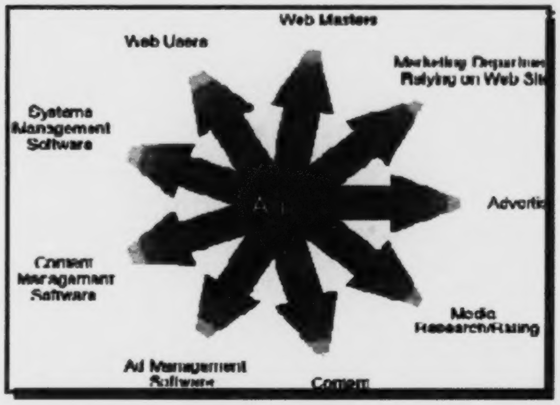
4) Advertising Subsidy growing dramatically

Any new medium of communication, commerce, education or entertainment needs to be funded by advertising. This is one rule of the past that still holds true in this current paradigm shift. Advertising dollars are the SEED funding or start-up funding of all new media delivery systems. The Web is no exception. Advertisers are starting to allocate money into advertising on the Web as the use of the Web increases, and as they see the potential for one-on-one marketing as a lucrative alternative to the mass-market advertising of TV and Radio.

Products from companies like Net Gravity, Focalink, Web.direct and DoubleClick make *management* of advertising in this new medium much more powerful and fully automated. The infrastructure participants of the last media revolution are also jumping into the Web, bringing with them their needed products and services. As metrics for evaluating revenue models and ROI (return on investment) emerge, companies like A. C. Nielsen are adapting their systems for rating media reach and value, so that advertisers can select the optimal places to advertise and programmers can charge equitable rates for advertising exposure.

All this leads to advertisers pouring large amounts of money into the Web to send their messages to their target audiences. Many of these audiences might no longer be reachable using the old media, especially among the young. Over the next 18 months advertising will increase its role in financing the Web media.

The Market



Aria delivers value to a broad range of companies and users relying on websites for their livelihood.

1) Web based Application Developers

It is only in knowing who (by category, not personally) is viewing a page that content management systems can dynamically generate the right page for the user and structure the appropriate links for that user. By knowing the psychographic category of a viewer, advertising management systems are able to dynamically rotate in the appropriate ad. It is only by knowing the network traffic patterns across an intricate set of sub-Web sites can a network management software product dynamically allocate resources in anticipation of future bottlenecks.

In addition, there are a number of companies such as Firefox and Webster providing corporations with tools for monitoring Web surfing and limiting access for employees only to sites useful in their jobs. For the Web to experience its full potential and for the application software products to fully benefit their target users, the users must be made aware of the detailed activity of the Web, both dynamically in real-time, and persistently over time.

2) Web Administrators

Web administrators and Web Masters need the following information about their sites:

- how to develop content customised to viewer
- navigational paths users take
- traffic patterns
- content update and refreshment

3) Marketing Organizations

Marketing organizations utilizing their website as a strategic vehicle to communicate information about their products, or positioning of their company, need to know the effectiveness of their site relative to their goals, and how well the site is doing in reaching the target audience. They need to know the source for the leads they are able to capture (i.e. the source link of users) and the detail of these sources matched against the quality. They need to know when content has gone stale or better yet, just before it goes stale. They need to measure the value of certain content in attracting the right type of user prospects. In general, the Web marketeer must know who is using the site and what is happening on the site in order to judge and improve the effectiveness of the website as a marketing tool.

4) Advertisers

Advertisers must be able to assess the value of certain advertisements on specific websites. An ad's value is measured in terms of breadth and depth. Breadth is the number of target customers reached with the ad. Depth is a measure of effectiveness of the ad with any one customer. Did the customer act on the message? Advertisers need to be able to quantify and qualify their spend on the Web.

5) Market Research Companies

Market research has been an "after the fact" task, whereby market research firms will contact people after they have made a purchase and try to learn more about why they purchased a particular item. Buying patterns and buying behavior are then extrapolated from these research results. Point of Sale systems capable of capturing information in somewhat more real-time sense, sending inventory reports back to suppliers and in general, improving the delivery chain process have improved the information chain. Wal Mart was a successful practitioner of POS programs. This system still does not begin to capture the entire experience with the consumer as it captures only the point at which the final purchase decision is made. POS says nothing about the 10 toasters the customer looked at and did not purchase; it tells the system only the one the customer did purchase. In the paradigm of the Web the potential to capture the complete customer picture for companies looking to market research now exists.

Conclusion

- Content, bandwidth, access and advertising will increase web traffic
- The market is comprised of principal users, including web based application developers, web administrators, marketing organizations, advertisers and market research companies

Chapter Two: System Requirements and Specifications

The requirements in the marketplace translate into a very tough computer science problem that is not easily solved via traditional approaches.

Having evaluated the direction of the Web marketplace and the genuine need for a scaleable, high performance Web activity recording, accessing and reporting system as requested by the marketplace, Andromedia has translated this request into the technical requirements driving our architecture design.

Technical Requirements

1) 100% Capture

The system must fulfill 100% data capture. If a stand-alone system is to fulfill the needs of the myriad of *applications* and satisfy the many classes of *users* needing reports. The system must capture ALL of the information available. The system must also accommodate the Web as it evolves, being able to capture *new types of data* as it emerges.

2) High Performance Capture and Retrieval

- Must keep up with the retrieval demands during peak load periods:

As a website's usage rises and falls, the system must be able to keep up without having to reconfigure the hardware on a dynamic basis. If an application is expecting real-time access to certain data, the system must be able to provide this access consistently.

- Must provide no loss of data due to overflow:

If the system is not supporting any real-time applications it will be acceptable to "catch up later", as long as the catch-up period is within the window in which reports on the data are being generated. In no case can the system get so overloaded that it is not accepting any more data and data is permanently being lost.

- Data must be available for retrieval immediately:

When real-time applications are depending on the data being made available immediately, the system must *maintain* this level of support. If an advertising management application is rotating local language versions of their ads depending on the domain reference of the user, that domain reference data must *always* be available.

3) Server Performance

The Web community is "performance elastic". If users run into performance problems and pages are not being served up quickly and consistently, they get frustrated and leave the site or the web altogether. This was apparent at the Internet Shopping Network in November 1995, when they experienced hardware performance problems for a one week period and their order rates dropped by 50%. The activity recording, accessing and reporting software must be configurable so as not to impact the primary web servers' performance.

4) Scaleability

The system must be capable of scaling intelligently to accommodate peaks and valleys in usage. This problem cannot be ideally solved by only adding hardware to accommodate the peak usage period, as it leaves the system grossly under-utilized during light usage times.

5) Efficiently utilize all Hardware Resources

The software must respect the cost of hardware and must make the most efficient use of all hardware components including CPU (including multiprocessing systems), memory, disk storage, I/O and communication bandwidth.

6) Standard and Open

- API for legacy applications to write data to the information system:

There must be a standard mechanism for enabling third party applications to write data into the database. There will be numerous third party applications generating data that must be aggregated with data that the activity accounting system captures directly from the web. The most obvious example here is a third party demographic analysis application that is compiling psychographic profiles for users off-line. This data will need to be matched against the user profile when the system detects the user is on a website.

- API for reading from the information system:

This is the primary API required by most third party applications, whether they are performing reporting and analysis functions or dynamic advertisement rotation. They will all need access to the information collected by the activity recording, accessing and reporting system. This access must be clean and structured in a way that supports the data model of the application. One API for all classes of applications will *not* work. Each "class" of applications will have a different set of requirements for viewing and accessing this information. This API must be extensible without forcing change in the existing API interface.

- Support all commonly used products and standards in the following areas:
  - Web Browsers: The user interface to the activity reporting, accessing and recording system must be web-based and therefore will be executed through a browser. Access to the activity reporting, accessing and recording system must be supported by the browsers representing 90% of the browser market.
- Web Servers (Software): The activity recording, accessing and reporting system must link into the most popular web server platforms (i.e. 90% of the web server market).
- Web Server Platforms (HW & OS): The system must run on enough platforms to cover 90% of the server hardware and OS market.
- Web Network Protocols: The system must work within a TCP/IP and HTTP protocol environment.
- Security methods.
- User identification standards.
- User authentication standards.
- Respect of privacy: The software must respect the personal preferences of the individual users for privacy. The user must have the choice of how much or how little they divulge about themselves. The user must also be assured that this information will only be used for specific purposes.

7) System robustness and availability

The system must be able to withstand failures of:

- commonly used hardware and software components. In the event of a failure, the system must restore itself smoothly and completely.
- The system must be easy to install, implement, manage and operate.

Chapter Three: Andromedia's Aria Web Recording, Accessing and Reporting Platform

Introduction

Aria dramatically improves the effectiveness of World Wide Web applications. It enhances the subjective experience of World Wide Web users and increases understanding of website activity for site owners.

Operating a website without Aria is equivalent to the National Weather Service tracking weather patterns without satellites.

Market requirements and technical requirements considered performance, system scaleability and object model extensibility as the most critical elements of a Web activity recording, accessing and reporting system. Andromedia's design goals were based on the assumption that the top ten web sites of today, measured in terms of activity and complexity of content, will represent the *average* web site on the Web in less than 18 months. With this assumption in mind, we constructed a system architecture that could handle peak traffic in excess of 10 million hits per day, can scale up and down smoothly making optimal use of all hardware components, and is completely open and flexible to accommodate the many unforeseen requirements for activity accounting information that may arise in the future.

Primary Design Choices

To respond to the general requirements in the market, Andromedia has made the following choices which serve as the skeleton of our system architecture and make up the basic architecture infrastructure in which we have implemented the Aria Web recording, accessing and reporting system.

1) Pure Object-oriented Approach to Building the System

There were four primary reasons for this decision:

- Crucial to enabling us to build a highly modularized system, whereby each module could be distributed and executed on multiple processors or multiple systems across a local or wide area network. It is this capability of distributing the processing that is one of the key enablers of our performance and scaleability advantage.
- Essential in supporting the level of extensibility the market requires. The object approach to schema design for the database uniquely enables additional API interfaces to be added without having to go back and re-build the entire database schema. This allows support to be added for new application interfaces without disturbing the API's already in place supporting other application classes.
- Over the lifetime of a product, object is the most productive software development environment available. More time is invested up front designing the object model and the overall architecture, but with the design done, new versions and extensions to existing products are published much faster than with conventional programming. This enables Andromedia to shorten product development cycles considerably, thereby getting newer products in the hands of our customers faster.
- Object is inherently the most suitable environment for building software systems that will reside on the WWW, as the Web itself is an object oriented environment.

Overall, the product Andromedia is building is expected to link inextricably to many other modules provided by companies other than Andromedia. Therefore, choosing an object paradigm for building our solution is critically important.

2) Multi-threading for Scaleability and High Performance

To capture and manage the volumes of data generated by websites today *and* project volume just 18 months from today, Andromedia must make full use of all available hardware in the network as well as hardware local to the servers running the network. Andromedia simply had no option of building single-threaded systems that only work serially on one process or one transaction at a time. It is required to load-balance the processing across hardware resources within a single system and across a network. Every available resource in the network must be considered a potential candidate to assist in balancing the load. Beyond simple course-grained "multi-process" architectures of the past 10 years, "threads" have become the leading way of achieving maximum performance out of available hardware. Threads are a finer grained approach to splitting up a task for processing. Thousands of threads can be combined to make up single processes. Without multi-threading the Andromedia system performance overall would be limited to the performance of a single thread path through the hardware and network platform. With multi-threading, the performance ceiling is lifted by at least a factor of 1,000.

3) Distributed Design

Although Object and Distributed often overlap when they are discussed, one does not necessarily imply the other. One can still have an object oriented system that does not distribute effectively. The Aria system architecture fully supports distributing the web activity recording, accessing and reporting information across multiple physical machines, multiple webservers, multiple geographies and multiple users. The integrity and control of this distributed information system is maintained in the Aria.store database module. The expected benefits of this approach are not merely increased performance and scaleability, but also enabling a higher level use of the global information base. Users will be able to easily compare their own results with those of participating companies in similar industries around the world by the click of a button. Individual companies will be able to easily aggregate data from websites running at remote divisions all over the world with a click of a button. If the underlying architecture does not support information being distributed logically as well as physically, trying to accomplish this after the fact is impossible.

4) Open, Standard, Plug and Play

Andromedia has chosen a design that completely adheres to the spirit of the Web in that we "plug in" without forcing any modifications to the "plugs" or "wall sockets". To initiate the capture of the "raw data", we do not require the web server software companies to make any modification to their software. We simply link directly in through their published API's as a **.so file** in Unix, or through a **dll** on NT. While we are processing the information for downstream or real-time use we are inside our own world applying our proprietary processing techniques (described later in this paper).

Once our Aria.recorder engine has completed its tasks the information is presented to the outside world through a published API that has been designed in cooperation with the leading web application companies looking to access this information. In addition, this API is completely extensible to accommodate the needs of application classes that will most certainly proliferate in the near future. The Andromedia API essentially defines the way an application will view the stored data either in real-time or on a persistent basis. These API's are called from either C or Perl function calls, with a Java interface planned for the second release.

The Aria Recording and Reporting System

The Aria system is a high performance and fully scaleable system for *recording* web activity data, *storing* this data for real-time and persistent access, *analysis* of this data for reporting purposes and *presentation* of the information to third party application consumers of this information. The Aria system is comprised of 5 discrete object modules linked through an architecture optimized for performance, scaleability and flexibility.

The following describes each module and the functions it performs, as well as the enabling technology and the resulting benefits.

Aria.monitor

Aria.monitor is a flexible shared library that runs dynamically on a Web Server and monitors all of the traffic between Web Clients and the subject Server. It supplies the underlying activity information that is essential to understanding how a Website is being utilized. A typical server log file does not capture all of the information that is captured by Aria.monitor. Standard log files do not capture "cookie" information or post information (HTML Form field information). They also cannot track transaction time, path transaction data and other desirable parameters.

Aria.monitor can run on multiple servers and over multiple sites. It can send messages to multiple recorders effectively creating duplicate sites for fault tolerance. A future Release will enable Aria.monitor to also write to its own logfiles as a fail-safe against dropped messages or recorder crashes. Since it is part of the site server it can accommodate changing volume requirements during peak surges of activity, at least as well as the server resources allow. It streams all of the client/server traffic information, including user cookie profiles, form field posted data, *and* all data ordinarily written to a log file to the Aria.recorder module. An open connection to Aria.recorder essentially maintains an open pipe to improve performance.

Aria.monitor records all hits on a page. A hit consists of each pointed element clicked on a page, and therefore, a single page can have multiple hits. Typically, each clicked image is treated as a hit, which is necessary to measure the effectiveness of multiple advertisements placed on a single page. Hot links to other pages are not treated as a hit until they are clicked on by a user.

Aria.monitor streams all of the Request, Response and Post data between the Web Client and the Server to the Aria.recorder. It also captures meta data with regard to the document being transferred between the Web Client and the Server.

Aria.monitor runs as a **.dll** on a Windows NT Server or as a **.so file** on a Unix server. It is automatically available when a server is booted, and is linked to the server through its API..

Captured Data

Aria.monitor is comprehensive with regard to the data it can capture including, if desired, all of the data from an HTML Form being transmitted between the Web Client and the Server. The following categories of data are captured:

- Client Request Data

The data captured by the Aria.monitor with regard to the Client Request includes the type of browser being used by the Web client, the authenticated user ID if the server requires it, the IP address of the client, any references to a previous page, the allowable response methods, and the acceptable data types that user can view.

- Client Post Data

In many website interactions, the user often enters field data into an HTML Form. Aria.monitor captures the body content of this data filled in by the user. This is key in certain applications, especially where the user is providing demographic data or a profile of his or her preferences. Applications can use this data to provide a personally customized view of their content to the user based on matching the user ID with previously stored profiles and/or preferences.

- Server Response Data

The data captured by Aria.monitor with regard to the server response includes location of source requested (could be another server), the hostname of the server, the date of last modification, any server defined directives, the expiration date of the data, and the server response code.

- Document Meta data

The meta data captured by Aria.monitor with regard to the document transferred includes the encoding type (e.g. base 64), the byte length of data transferred, the type of data transferred (e.g. GIF, cgi script), and the body content of the message.

- Cookie Data and user profile data

Cookie data refers to the path taken by a user during a prior visit. Aria.monitor can capture this data so that the server is aware of prior paths and can automatically customize the users' interactions based on patterns and preferences previously exhibited by the user. If the server does not support cookies the Aria.monitor creates one for the user. In addition, under the control of the administrator the Aria.monitor product can allow user authentication files or other pseudo-cookie alternatives to be used in place of cookies to maintain the user profile.

All of the above data is streamed in real-time to the Aria.recorder engine, which intelligently and dynamically transforms this data stream into *persistent objects* for storage and analysis purposes. In addition, as the website evolves over time and more or different data is passed between the client and the server, Aria.monitor is set up to capture that as well. Aria.monitor is designed to capture all information it sees, and let the Aria.recorder sort out what information is needed and what information is not. It also takes up virtually no server resources to run, using less than 50K in disk space.

Aria.recorder

Aria.recorder is an extremely robust data management engine that accepts the data stream provided by the Aria.monitor, which in turn is capturing all Web Client and Web Server interactions. Aria.recorder can run on the main Web Server or on a distributed server on a local or wide area network.

Aria.recorder is designed for high performance and scaleability to manage efficient processing during peak surges of activity. A single Aria.recorder engine can accept data streams from multiple Web Servers, each running Aria.monitor. It is a high performance daemon that is fully scaleable to meet changing volume requirements. During times of peak activity - if hardware is not capable of handling the load, the recorder creates optimized backlog files. These files store the update data in an optimized way, such that when usage/loading relaxes the backlog data can be employed very quickly to update the database.

Configuration of the recorder can be optimized so that it performs optimally on any hardware configuration. Larger RAM and disk storage or SMP multi-process servers can execute the recorder transactions very rapidly, consistent with their superior capabilities for load handling.

The granularity of data being captured by Aria.recorder is configurable by the site administrator depending on traffic volumes and specific application needs. Such a filtering of the data is necessary to avoid persistent storage of data that is not useful. A high volume site can easily generate up to 400 megabytes per hour, with exponential growth expected for the future.

Aria.recorder parses the incoming data stream, labels each element for the appropriate object, and creates the database objects for persistent storage.

Thread Based Architecture

Aria.recorder uses a multi-threaded architecture to make efficient use of processor resources and is extensible depending on specific site and application needs. It can spawn new threads to keep up with increasing data loads. The number of threads that can be running simultaneously is limited only by the available memory and swap space on the disk.

The Aria.recorder is also capable of spawning completely new "virtual" Aria.recorders that each answer on different communication ports. This is very useful to administrators that may be hosting multiple websites and need to guarantee access bandwidth to each site as well as manage the activity accounting information from one site separately from that of another site.

Aria.recorder may run as a single process, but the number of threads used by the engine are user configurable and can be changed dynamically. It can support hundreds of threads which are only limited by the memory and swap space of the specific hardware in place.

There are six basic classes of threads:

- Accept
- Buffer
- Update
- Transaction
- Scheduler

- Reaper

Web administrators can allocate the maximum number of threads to each class, depending on specific hardware limitations to optimize system resources.

The Accept thread grabs the data streaming from the Aria.monitor and throws the data onto a connection stack. *The Buffer thread* then grabs the data from the connection stack and places it into a buffer stack, releasing the *Accept thread*. The Buffer stack can expand in memory to accommodate peak loads, so the data can be processed later as more system resources become available during off-peak hours. If Andromedia's Aria.reporter is creating real-time reports, then additional hardware can be added at the site to minimize buffering and accommodate the immediate processing of all captured data. (A future Release will call an API to read the latest updates.)

The Transaction thread parses the buffered data through a series of functions that are extensible based on specific site and/or application needs. The unique architecture of this extensible thread provides tremendous flexibility for configuring Aria.recorder. The functions used by this thread are customizable for specific site and/or application needs, and can be easily modified as requirements change over time. All real-time data needs are controlled by this thread. Data elements that have been tagged for real-time access are fed directly into memory and made immediately available to the application counting on it.

The Scheduler thread is used as a filter to purge information on a periodic basis. It is used for sampling purposes and for creating reports based on persistent data that has been previously captured. *The Reaper thread* is used for aging the data and clearing disk space by deleting obsolete data based on user specified rules. *Aggregate* information on the purged data is still maintained even when the content data has been purged. This ensures that no relevant statistical information is ever lost.

Data Objects Created

Aria.recorder creates the following objects for persistent storage by Andromedia's Aria.store module:

- WebUser
- WebUserAggregate
- WebServer
- CategoryIntersection
- WebObject
- WebObjectAggregate

The WebUserAggregate and WebObjectAggregate, along with WebServer, generally stabilize in size for any given server site. The WebUser object identifies the user, the WebServer object identifies the server and the WebObject identifies the content (such as a HTML page, Java applet etc.). These objects then expand only when new users, servers and/or content types are encountered.

While the start time of a visit is always captured, due to the stateless nature of the HTTP protocol, current web technology does not inform the server when a user has *finished* a visit and moved onto another page. Aria.recorder, therefore, allows web administrators to select a default "Time Out" to be used for ending a particular visit. By consistently comparing visits with similar time outs, it is possible to approximate the end point for the visit.

After creating the objects Aria.recorder passes these to Aria.store for writing to the database.

Categories

The Aria.recorder daemon uses categories to translate obscure names or words into meaningful terms, and creates groupings of field values with similar traits. These categories (see "System Administration Tools) are user configurable. They are created in files of *regular expressions* which can be used on any of the four fields - *URI, Referrer, User Agent* or *Domain.*

Aria.store

Aria.store manages the real-time and persistent storage of objects created by Andromedia's Aria.recorder. These objects are cached before being written to disk, to enable applications requiring real-time access to the data the immediate access they are depending on. At peak load times the data can be buffered for later processing during off-peak hours when more system resources become available. If Andromedia's Aria.reporter is being used to create real-time reports, then additional hardware can be added at the site to minimize buffering and accommodate the immediate processing of all captured data.

Aria.store supports multiple databases and sets up the Schema for the selected database. Single databases can also be partitioned across multiple machines, either for added performance or additional resource control.

Aria.api

Aria.api is the Application Programming Interface that allows access to the stored objects in the Aria.store database. This API permits the Aria environment to be completely open and extensible. The API defines how standard scripting languages such as Perl, Java, Tcl and C++ can access the data objects created by the Aria.recorder engine.

The first API provided by Andromedia interfaces Perl5 to the ObjectStore database. This allows developers to write Perl5 scripts and add Perl5 modules that are accessible to users from web browsers. An application server will allow the Perl5 interpreter to be wrapped in an executable file and be aware of any Andromedia or user defined extensions. This avoids the need for the interpreter to be initialized and eliminates CGI forking, resulting in optimized execution speed.

Third parties have the ability to access stored data objects and data structures through the Aria.api on both a real-time and persistent basis. Developers can build powerful new applications or interface existing applications to the web activity data through this API. This very same API is used by Andromedia's Aria.reporter to analyze the stored data and create reports.

API Object Classes

The Web object model is organized into three fundamental classes; the users, their activities and the objects they manipulate. Users and objects are relatively finite, but their interactions are enormous. The object classes exposed through the API are Users, Servers, Pages, Visits, Statistics, Queries, Graphs and Reports. The first four classes can be monitored on a real-time basis. In a future Release the user can view aggregate data from these objects on a monitoring dashboard that displays the on-line information in real-time.

The user object represents a *specific user* of the web site. It basically saves two kinds of information about the user. First, it saves *account* information which includes a site unique user id, when the user first visited, and when the account expired (best approximation of when the user left the site). The User object also saves a *persona* which is how the user would describe himself via a form or other information they control and make publicly available to web sites.

The server object maintains information about the type and configuration of the server. It also tracks server statistics.

The page object tracks the properties of each *web object* in the system. Objects that don't get hits effectively don't exist. Page objects get created on the first hit. The FirstClickTime is the time the current revision of the page object is used. The stats object gets updated regularly, while page object attributes will also be updated occasionally, signaling that content has been revised. In this case, a *new* page object is created with the same URL, a new FirstClickTime, the new revision, along with the updated attributes are also captured. This will allow queries across the same URL after content changes, and still retain a history of the URL attribute changes mapped against the activity.

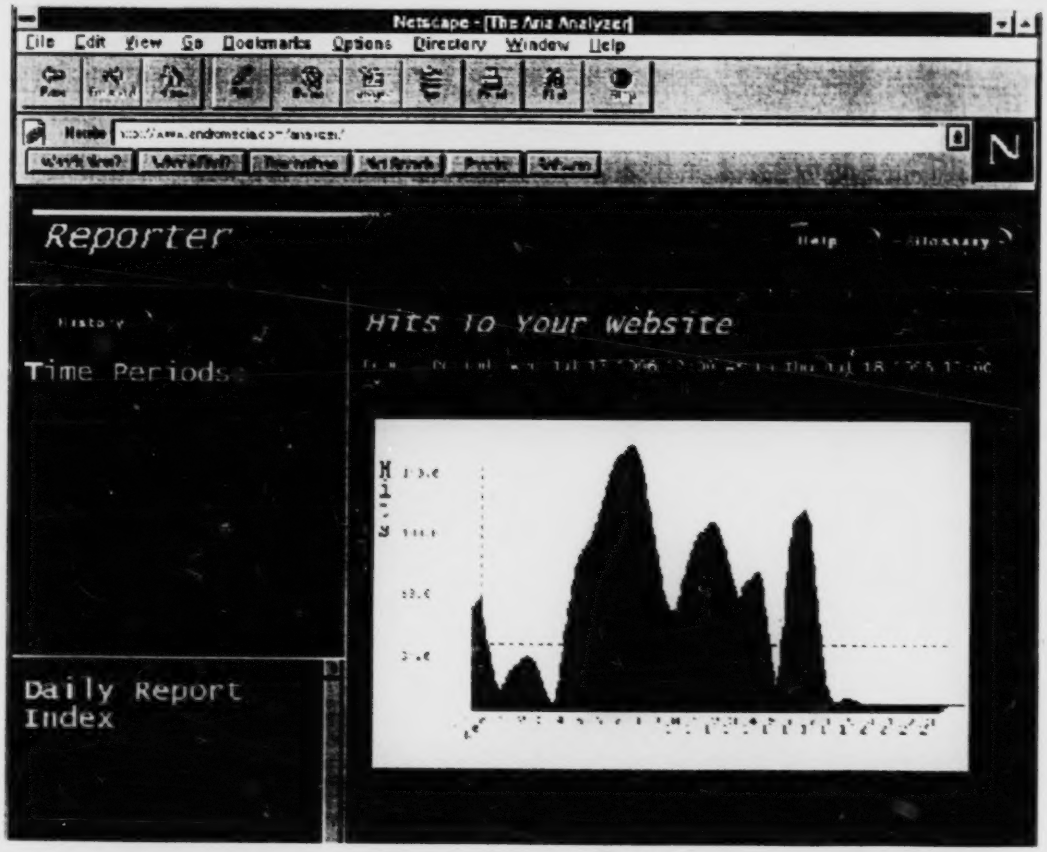
The Stats class object is a generic object class reserved for saving statistics over time. Specific stats are aggregated over a time slice based on time granularity settings. Aggregation can occur hourly, daily, weekly, monthly or as a single total. Finer granularity intervals can be set but take up considerably more disk space while providing a more precise scan of the trends in activity.

The Queries classes are used in conjunction with the Graphs and Reports classes for ad hoc querying of the stored data.

Aria.reporter

The Aria.reporter creates an intuitive graphical user interface and solves two very difficult problems inherent to the Web. The first problem derives from the roots of the Web's client/server architecture. The mechanism of communication between the browser and the web server is a series of request and response message passing. This back and forth message passing is *stateless* -- the reality of a *browser session,* having a beginning and an end, does not exist. Unfortunately, much of the site activity information that is relevant to Aria users requires a website visitor having a *session* (visit) that begins and ends. The Aria.reporter handles this problem of stasis by allowing a system configurable setting that *creates* virtual state, through the implementation of a visitor timeout. What this means is a new visit session begins when a browser arrives at the site (which *can* be detected), continues while the browser is active in the site, and ends when a new visit starts. All the important information that webmasters, marketing/sales personnel and customers need to analyze the web site fall out of the introduction to this concept of *state*.

The second problem is one of *visualization* of data. How does one visualize information about website visits, paths taken, browsers used etc.. Other companies engaged in the web tracking market have resorted to the most practical route by presenting the data in tabular form.


Netscape - [The Aria Analyzer]
File Edit View Go Bookmarks Options Directory Window Help
N
Reporter
Help
Glossary
History
Time Periods
Hits To Your Website
Hits
Daily Report Index

BEST COPY AVAILABLE

The Reporter user interface provides the viewer a simple, powerful and intuitive way to visualise the activity on the website.

While this approach makes sense it is not necessarily the most intuitive way for an Aria user to view the data. For example if one is looking at the browsers who hit a website, is it intuitive for a user to see numbers like

Browser	Number
Mosaic	345984
Netscape 1.1	235987
Netscape 2.0	123004
Spyglass	320008

Moreover, is the *daily* aggregation of the data about browsers a useful statistic? Is that the correct time granularity? The traffic on the web is inherently *dynamic,* and the web site itself is constantly changing in terms of content and interest level. The aria.reporter solves the problem of intuitively presenting the data by displaying the information in *graphical* form, using pie graphs, bar charts, histograms and time-series wave graphs. Moreover, by presenting the graphs in three dimensional perspective, end users can view the data spatially and measure relative size. The idea of measuring data multidimensionally, i.e. relative to other data is a key implementation of the aria.reporter; website activity data only makes sense in relative terms, not in absolute terms.

Another key implementation of the aria.reporter is one of presenting the data through relative time graphs, so that users can see not only how the data compares relative to one another (slice and dice), but also how the data changes over various time periods (drill down). Using graphs in addition to tabular reports, aria.reporter is able to make more sense of the data.

The user interface of the aria.reporter allows decision makers to access the information quickly. Web architecture is inherently object-oriented. Fundamentally the web is comprised of many distributed nodes (objects) sending requests and responses to each other. The web's dynamic nature requires a more sophisticated logging facility that can adapt over time. Persistent storage must be able to handle evolving schemas to keep up with people's changing perceptions. Andromedia stores this information in an *object database* which makes the retrieval of information easy and flexible. Through an HTML interface, end users select the filtering mechanism of the objects they are interested in (i.e. visits, statistics, bandwidth etc.) and the information is pulled out of the database. The flexibility derives from the fact that through this HTML user interface, users create graphical reports on the objects that they want by filtering on the encapsulated data members.

Aria.reporter takes advantage of the click-through capability of the web. In a typical aggregation report a user might see an ordered list of bars related to a topic, and then by clicking on a particular bar, the user can expand on that bar to see the detailed time-series view of the data related to that bar.

Aria.reporter has in effect modeled its implementation on the architecture of the Web itself. The Web produces object structured output that is generated in time series. Through spatial graphs, aria.reporter analyzes the state of the web site, which is then captured and presented to the end user.

Administration Tools

Andromedia provides (currently under construction) a comprehensive set of administration tools that can be used to configure the Aria product and to manage the activity data for the site. For example Website administrators can configure the installation of Aria.monitor and Aria.recorder to be resident on the same server or on different servers. These tools also allow administrators to filter out data that is irrelevant based on the specific needs of the site and/or the application.

By using these Administration Tools, Aria.recorder can be configured for the type of data of interest to the site administrator and/or the application. This is especially important for popular sites which may be experiencing over a million hits per day and can, therefore, generate large volumes of user interaction data very rapidly. By selective use of *categories* the recorder can translate obscure names into meaningful terms, and create "groupings" (aggregation) of field values with similar traits.

Currently the following fields can be categorized:

- URI - Uniform Resource Locator
- Referrer - Link that leads to the current page
- User Agent - Browser type
- Domain - Domain that the visitor comes from

Another powerful facility of the Administration Tools is the ability to compact aged data statistics. Aged data becomes less relevant and can be safely compacted through this facility. Administrators can define rules, for example, maintain today's data on an hourly basis, prior seven days data on a daily basis, and the previous year's data on a monthly basis. Older data can also be moved to off-line storage for archiving purposes.

Developer's Toolkit

Andromedia provides developers with a toolkit that enables them to extend the Aria model to incorporate new objects that may be required to support new applications. The toolkit consists of object classes that link the Aria.store database (e.g. ObjectStore) with a programming language (e.g. Perl5) to a web client (e.g. Netscape). The Developer's Toolkit can be also used to optimize the performance and reliability of the Aria system.

Summary/Conclusion

Table of Technical Requirements vs System Features

Technical Requirements	Architecture Response	System Features
100% data capture	Aria.monitor	No loss of data
High Performance	Aria.recorder Threads	Scaleability
Zero performance impact on main web server	Distributed architecture	Optimize use of all hardware components
Real-time access to information	Aria.store Object	Persistent information management
Adherence to standards	Aria.api c, Perl, C++, Java	Open access to information base
Robustness	Aria.reporter GUI	Ease of Use

The above table summarizes the technical requirements of the market mapped against the architecture features we chose to implement.

To summarize:

1. The anticipated future use of the web will see exponential growth relative to today's average use. Activity record data will be generated, on average, at a pace of over 1 GB per day and will see even greater bursts. The underlying system capturing and feeding this information to applications must be able to manage this growth.

1. The Aria Web Recording and Reporting System was designed to address the high performance and scaleability issues of the Web marketplace. By adopting an object oriented approach to the architecture design, exclusive use of multi-threading, and the ability to fully distribute the various modules across local area or wide area server platforms, Andromedia can fully satisfy the performance and scaleability needs of the marketplace.

Automating Hypermedia Course Creation and Maintenance

MAJ Curtis A. Carver
LTC Clark Ray
carver@eecs1.eecs.usma.edu ray@eecs1.eecs.usma.edu
Department of Electrical Engineering and Computer Science
United States Military Academy
West Point, NY 10996

Abstract: While many tools are exist for the development of individual Hypertext Markup Language (HTML) pages, few exist for the development of complete hypermedia courses. The HTML Course Creator (HCC) allows instructors who are not HTML experts to rapidly develop and easily maintain consistent, distributed hypermedia courses and associated digital libraries. Using a simple point-and-click environment, instructors can add a variety of media types without knowing a single HTML tag. The tool acts on user input to create a hierarchical course structure and all of the related HTML documents. The course can be tailored to specific styles based on templates that are consistently applied to all course documents. Furthermore, the software maintains a complete course-wide digital library of all media used within the course by location, media type, media anchor, and lesson. Using the HCC, instructors from a variety of academic backgrounds can create and maintain networked, hypermedia courses accessible over the World Wide Web.

Introduction

Like many other institutions, the United States Military Academy (USMA) has developed hypermedia-based courses using Hypertext Markup Language (HTML) to exploit the World Wide Web (WWW) as a means of enhancing learning outside the classroom. At USMA, development of these course materials has until recently been primarily in the domain of computer scientist and information technologist despite an extensive network infrastructure reaching each student's personal computer. The initial hypermedia course developed at USMA, CS383 Computer Systems, has grown to over 1.2 GB of course material as depicted in [Table 1].

• Lesson Objectives, Note-taking Guides, and Electronic Slideshows for every Lesson	• 143 Audio, 63 Graphic and 57 Digital Movie Files
• 300 pages of course hypertext with 678 terms with pop-up definitions and 600 terms that can be searched on.	• An animated virtual for benchmarking typical computer configurations.
• An adaptive testing system based on Common Gateway Interface (CGI) forms with over 250 questions, hints, and color-coded reporting based on lesson objective and depth of learning according to Bloom's Taxonomy.	• An adaptive hypermedia CGI interface based on Felder's Learning Style Model that tailors the presentation of course material to the learning style of the user.
• A course legacy system with over 275 student papers and slideshows from previous semesters supporting a paperless student submission and grading system.	• An ISMAP-based user interface of over 1500 links interconnecting all course material.

Table 1: Hypermedia CS383 Components

Despite very positive results using hypermedia courseware [Carver & Howard 1995a] and similar efforts at other institutions, many departments at USMA did not immediately move to exploit the WWW for educational hypermedia. The most significant barrier to persuading instructors to invest the enormous amounts of time required to produce and maintain hypermedia materials for courses is the well-justified concern that there may not be a suitable return on this investment. Concerns also exist regarding student acceptance and learning outcomes, and technological risk-factors associated with hypermedia courseware such as architectural standard isolation, network access bottlenecks, and tool longevity. A mechanism was needed to reduce the amount of time required for hypermedia course development and maintenance. If time requirements were significantly reduced and the methodology was simple yet powerful, then there might be more interest in developing hypermedia courseware.

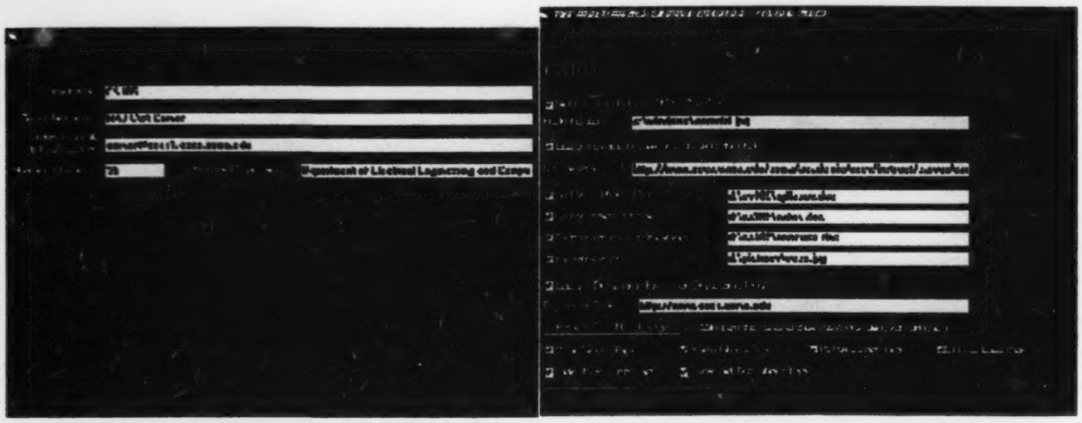


Figure 1: HTML Course Creator Key Information and Main Course Interface

There has also been widespread interest in interconnecting the courseware developed across departmental boundaries so that *hyperdisciplinary* courseware - hypermedia course material from different disciplines - integrated seamlessly through a common interface with rich cross-linking and sharing of resources. Instead of students learning cross products from the mathematics curriculum and then again in engineering courses using separate materials, there could be explicit links between all courses using cross products so that teachers and students could reuse and extend on the course material from different departments. One of the first steps in developing hyperdisciplinary courseware is a tool that standardizes, to a useful degree, the hypermedia course interface. Since the maintenance of substantial quantities of hyperdisciplinary material would be extremely difficult without automated support, what was really needed was a one or more complementary tools with consistent user interfaces that combined hypermedia course development and course maintenance into a single consolidated package. Unfortunately, no such package was known to exist.

Previous Work

A variety of hypermedia tools have been developed such as GETMAS [Wong et al. 1996a], HM-Card [Mayrhofer et al. 1996a], Hypercourseware [Siviter and Brown 1992a], Hypertactics [Mulhauser 1992a], ISAAC [McAleese and Ching 1993a], MALL [Tanaka et al. 1996a], Metaplant [Hedberg et al. 1996a], NEAT [Mayer et al. 1993a, Muldner et al. 1996a], and Nestor [Jonassen and Harris 1991a]. A variety of public domain, shareware, and commercial software tools have been distributed for the development of HTML documents such as *HTML Writer*, *HTML Assistant*, *HotDog Pro*, and *Internet Assistant*. None of these tools provide support for course digital libraries, point-and-click environment for adding MS PowerPoint and MS Word for Windows documents with no required knowledge of HTML, or producing courseware with consistent student interface which are essential for hyperdisciplinary applications. In this regard, this paper describes a tool that is both novel and significantly different from previous efforts in this field.

The Solution - the HTML Course Creator

A group of our students developed a software tool, the HTML Course Creator (HCC), that addressed these problems. This tool facilitates the development and maintenance of hypermedia courseware. Furthermore, the software maintains a complete course-wide digital library of all media used within the course by location, media type, media anchor, copyright source, copyright release, and lesson. Using this tool, the creation and maintenance of large, richly interlinked, hypermedia courses has been both greatly

BEST COPY AVAILABLE

simplified and standardized.

Creating a course using the HCC consists of three steps. In *Setting up the Environment*, the instructor enters general course information [See Fig.1] and sets the directory that the HCC will compile into. *During Adding Media*, the instructor adds different media resources to be used in the lessons [See Fig. 2 and 3]. As media is added to the course, it automatically becomes part of the course digital library [See Fig. 3]. In *Course Compilation*, the HCC takes the information entered and generates appropriate HTML code in the compilation directory

The first step in creating a new course is entering the course information [see Fig. 1] and setting up the compilation directories. The information from the course information and main page data entry screens are combined to form the HTML course structure and the main course page. Instructors simply fill-in-the-blanks or indicate the file that they want included. If an option is left blank, a link to the option is not generated when the course is compiled. Instructors can choose between 1 and 40 lessons to include in the course. The program will use this information during compilation to build the course. The compilation directory is the directory where the HCC will create the HTML version of the course. It is normally a sub-directory on the departmental WWW server.

Figure 2: Course Creator Course and Lesson Interface

After selecting a lesson, the HCC allows the faculty member to select various media elements stored in institution- or department-wide servers and include them on the course home page, in the course digital libraries, and in particular lessons [see Fig. 3]. Instructors need only "point-and-click" on the media resource and then add a text description of the media resource.

The HCC intentionally supports incremental development of media. Due to the time required to generate hypermedia courses, few instructors have the time to generate a complete course in a single semester. Instead, most hypermedia courses are generated over several semesters and constantly refined. As instructors generate media, they can add the media to the course. Instructors can start with as little as text-based lesson notes for each lesson and then incrementally develop the course until it includes text, slideshows, audio files, graphics files, digital movies, and links to other WWW sites.

During compilation, all of the files indicated by the instructor are copied to the compilation directory according to a predefined directory and file structure. The HCC will then dynamically build links to all of the media files and insert the information entered by the instructor according to predefined HTML templates. These HTML templates support the ladder model of hypermedia courseware where media is linked both by lesson and by media type [Haga & Nishino 1995a]. This provides multiple paths through the course material that the user can select based on their preference. Currently the program supports two templates although additional templates can be added to provide a new HTML structure.

When a course is compiled, the HTML code is generated according to the template selected. The template defines the appearance of the HTML pages. Currently, two templates have been implemented . The first template compiles the HTML code so that course information is connected through simple

HTML links. The second template provides inline graphics in addition to the first approach. Both approaches provide a simple, yet orthogonal user interface. Once students have learned how to use the interface of one course developed by the HCC, they will have learned the interface into any other course developed by the HCC.

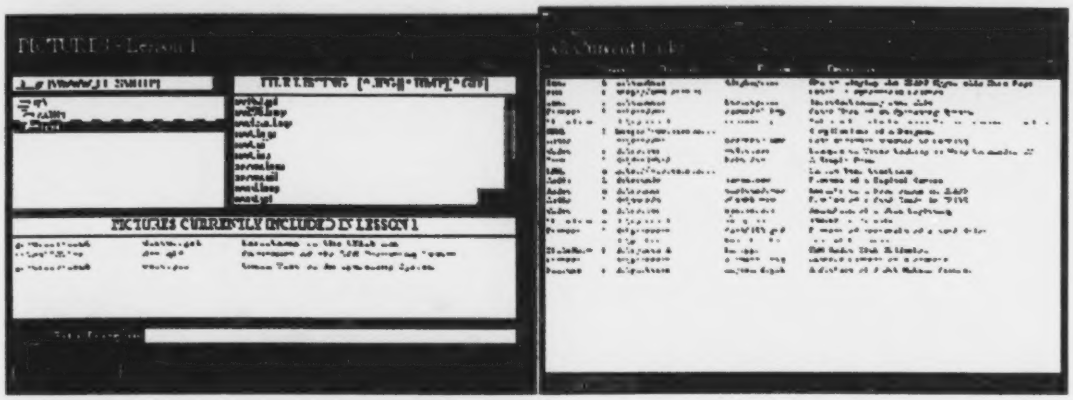


Figure 3: Add Graphic Interface and Digital Library Interface

The maintenance of a course built with the HCC is easy. As an instructor builds the course, the HCC maintains a project file with all of the information previously entered . Professors can easy copy and paste media elements between lessons or view media usage through the HCC digital library. Work is currently ongoing to provide drag-and-drop lesson reordering.

The maintenance of networked hypermedia, not the rapid generation of courseware, is the greater concern for course developers. The HCC simplifies tracking what media is used in what lesson and associates a plain text description of media with the media file. Media can be tracked by all of the different types [see Fig. 3] or it can be tracked by media type. This provides a simple yet effective mechanism for dynamically tracking what media is used in what lesson.

Advantages and Disadvantages of the Course Creator

The HCC provides six major advantages: it is a course-based, not page-based HTML development tool; simple professor interface; incremental course development with the ability to link in shared and personal media elements; consistent student interface and hyperdisciplinary course support; production of a course digital library; and modest machine and network resources required for effective use. Unlike other tools, the HCC supports a course-based approach to the development of hypermedia courseware instead of a file-based approach. This object-based approach more closely models other, more traditional methods of course development, and as such is easier and more natural to use.

The simple point-and-click professor interface complements the course-based approach by supporting instructor development of hypermedia courseware without a need to know HTML. This allows professors to focus on the content of the media added and not on HTML. It is believed that this will greatly reduce the amount of time required to create a course. Incremental course development allows instructors the flexibility to develop as much or as little of the hypermedia course as time or desire permits. Few professors have the time to monolithically develop gigabytes of course material. Incremental development is often the only viable approach for hypermedia courseware. Instructors may begin with a simple course outline and lesson objectives and as time and other resources permit, develop and include other media resources such as lesson notes, slideshows, sound files, graphics, movies, and links to other information sources. Furthermore, the media elements developed can be easily shared with other courses due to the consistent directory structure produced by the HCC.

The HCC provides a consistent student interface. Once a student learns how to use the interface in one course, the student knows how to use the interface in any other course. This allows for seamless

BEST COPY AVAILABLE

cross-linking between courses and integration of courses across departmental boundaries and eventually, hyperdisciplinary course support. Without a tool such as the HCC, hyperdisciplinary courseware is essentially impracticable.

The HCC requires moderate machine and network resources. It has been successfully used on relatively modest platforms (slow 80486 machines with 8MB of RAM, 500MB local hard drives, and standard shared 10 Mbps Ethernet network connection). It outputs standard V1.0 HTML, so a wide variety of Web browsers can be used to access the courseware created. Network file server and Web server requirements will, of course, still vary with the amount of material being used and the number of simultaneous student accesses anticipated.

The HCC has three major disadvantages: lack of flexibility in course design, no support for Common Gateway Interface forms, and no support of incremental compilation. The lack of flexibility limits course design to one of the two templates available in the HCC. Instructors cannot individualize their course using other HTML editors as the tool will overwrite these changes the next time the course is compiled. Likewise, there is no support for CGI forms, Java scripts, frames, or some of the other emerging features of HTML. While these are significant disadvantages, hyperdisciplinary courseware will not be possible without some standard interface. Furthermore, because the HCC was designed to dramatically reduce the amount of time required to create an HTML-based course, inclusion of these features would not facilitate rapid course development. A powerful, yet simple tool was needed.

Lack of incremental compilation remains a limitation of the current system. When the HCC recompiles a course, it recopies all of the media elements to the HTML directory and recreates all of the code necessary to link the course together. This overwrites any previous code or changes made to that code from other HTML editors. Clearly, it would be more cost efficient in terms of time and network bandwidth to only changes those portions of the course that had been altered. It is the intent of the authors to add this feature to the software over time.

Longevity of the HTML Course Creator

A problem with any locally developed software artifact lies in finding viable ways to maintain and extend it over a sustained period of time. Even reliance on commercial products does not alleviate all of these concerns. Few instructors will desire to commit to a tool whose problems are unfixable and which can not evolve to match developments in the problem domain it purports to address. Currently, much of the coding and testing work which has and will go into evolving the HCC is undertaken by the annual crop of senior students undertaking software design group projects; strategic direction and oversight remain with knowledgeable faculty members. In this mechanism is a fortuitous linkage between the desire of most groups to create a new artifact reflecting their own creativity and development tool preferences, and the need to substantially modify hypermedia courseware tools every year to keep up with changing standards and instructor requirements. A prior year's working tool becomes the living specification of the existing baseline requirement (adjusted for known bugs) and the basis for prototyping the current year's extensions and modifications. The long term viability of this approach to evolving the HCC remains to be seen, but it has so far proved reasonably effective and admirably (though not optimally) economizing of faculty time.

The HCC is currently being used in several departments at USMA as the basis for hypermedia course creation. Faculty members as well as the department leadership from several departments have been enthusiastic. One department has decided that all hypermedia courseware within the department will be developed using the HCC. Validation of exactly how much time is saved and the degree of acceptance among professors who had previously been unwillingly to exploit the WWW remains an ongoing research topic.

Summary and Future Research

The HCC simplifies the creation of hypermedia courseware. Through a simple point-and-click interface, professors can rapidly and easily build hypermedia courses. Construction of the hypermedia course is focused on lesson construction and not page construction as other tools require. No knowledge of HTML is required. Maintenance of a HTML-based course is likewise simplified with digital library support and media tracking. Requiring only modest platform support, the HCC supports the development of hyperdisciplinary courseware. It represents a first step in providing hypermedia courseware design software that anyone and everyone could use.

Future research into the HCC is focused on enhancement and assessment. Incremental compilation, support for the "chunking" of related lessons, color-coding material that must be updated on a semester basis, and attempts to improve the speed and functionality of software are the focus of the enhancement effort. The assessment effort is focused on validating the amount of time saved and the degree of acceptance among professors who had previously been unwillingly to exploit the WWW. Based on experiences with the tool, it is believed that HCC will have a significant impact on both.

References

[Carver & Howard 1995a] Carver, C. A. and R.A. Howard (1995). "An Assessment of Networked Multimedia and Hypermedia" *Proceedings of the 1995 IEEE/ASEE Frontiers in Education Conference*. Atlanta, GA. 2c5.17- 2c5.21.

[Haga & Nishino 1995a] Haga, H. and M. Nishino (1995). "Guidelines for Designing Hypermedia Teaching Materials". Proceedings of *ED-MEDIA 95: World Conference on Educational Multimedia and Hypermedia.* Graz, Austria. 283-288.

[Hedberg et al. 1996a] Hedberg, John, Barry Harper, Robert Wright, and Grant Farr. "An Authoring Metaphor to Match Constructivist Theory" *Proceedings of ED-MEDIA 96: World Conference of Educational Hypermedia and Multimedia.* Boston, MA. 776.

[Jonassen and Harris 1991a] Jonassen D. H. and D. Harris (1991). "Analyzing and selecting instructional strategies and tactics". *Performance Improvement Quarterly*, 4.

[Mayer et al. 1993a] Mayer, S., T. Muldner, and C. Unger (1993). "NEAT: An Integrated Authoring Environment Based upon Toolbook". *Proceedings of EDMEDIA 93: World Conference on Educational Multimedia and Hypermedia.* 332-339.

[Mayrhofer, V et al. 1996a] Mayrhofer, V., Nick Scherbakov, and Keith Andrews. "HM-Card: A New Approach to Courseware Production" *Proceedings of ED-MEDIA 96: World Conference of Educational Hypermedia and Multimedia.* Boston, MA. 433-438.

[McAleese and Ching 1993a] McAleese, R. and L. Ching (1993). "An instructional design advisor for computer-based learning: ISAAC". *Proceedings of EDMEDIA 93: World Conference on Educational Multimedia and Hypermedia.* 68-75.

[Muldner et al. 1996a] Muldner, T., Kasia Muldner, and Christine Marie van Veen. "The Design and Evolution of an Authoring Environment and its Applications" *Proceeding of ED-MEDIA 96: World Conference of Educational Hypermedia and Multimedia.* Boston, MA. 503-508.

[Mulhauser 1992a] Mulhauser, M. (1992). "Hypermedia and Navigation as a Basis for Authoring/Learning Environments" *Educational Multimedia and Hypermedia*, 1, 51-64.

[Siviter and Brown 1992a] Siviter, D. and K. Brown (1992). "Hypercourseware". *Computers in Education.* 18. 163-170.

[Tanaka et al. 1996a] Tanaka, Koji, Soichiro Iga, and Michiaki Yasumura. "The Development and the Evaluation of the Multimedia Assisted Language Learning Environment: MALL" *Proceedings of ED-MEDIA 96: World Conference of Educational Hypermedia and Multimedia.* Boston, MA. 661-666.

[Wong et al. 1996a] Wong, Wing-Kwong, Tak-Wai Chan, Yao-Sheng Cheng, and Shih-Shen Penh. "A Multimedia Authoring System for Building Intelligent Learning Systems" *Proceedings of ED-MEDIA*

96: World Conference of Educational Hypermedia and Multimedia. Boston, MA. 708-713.

Methodological Issues in Studying and Supporting Awareness on the World Wide Web

Lee Li-Jen Chen and Brian R. Gaines

Knowledge Science Institute
University of Calgary
Calgary, Alberta, Canada T2N-1N4
{lchen, gaines}@cpsc.ucalgary.ca
http://ksi.cpsc.ucalgary.ca/KSI/

Abstract: The World Wide Web has grown very rapidly to become a major resource supporting collaborative activities in a wide range of groups, disciplines and communities. However, the growth of the web creates problems of information overload and of maintaining awareness of activities at other sites relevant to one's own tasks. This article develops a methodological framework for studying and supporting awareness on the web; describes Chrono, a tool for supporting awareness of changes at another site; and uses the framework to classify Chrono and related tools, and to clarify the human factors design issues involved.

INTRODUCTION

The exponential growth of the World Wide Web (WWW) and the growing availability of collaborative tools and services on the Internet have facilitated innovated knowledge creation/dissemination infrastructures, such as: electronic libraries, digital journals, resource discovery environments, distributed co-authoring systems and virtual scientific communities [Schatz91]. Collectively, these Internet infrastructures have become integral parts of an emerging information system supporting collaboration in distributed scientific/research communities.

World Wide Web was originally conceived and developed at CERN for the purpose of assisting and facilitating collaborative interactions among high energy physicists, working at various institutions in different countries, to conduct joint research projects [Berners-Lee94]. Since 1993, it has diffused at a phenomenal rate and gradually has subsumed various popular Internet services such as: USENET Newsgroup, Electronic Mail, etc. The original charter of the web is summarized in the following quote: *The World Wide Web was developed to be a pool of human knowledge, which would allow collaborators in remote sites to share their ideas and all aspects of a common project* [Berners-Lee94].

However, the growth of the web, while creating a rich new resource, also creates problems of information overload. The management of the diffuse communities collaborating through the web raises human factors issues going beyond those of the coordination of smaller, goal-directed groups with well-defined roles and tasks. For example, what are the responsibilities of information providers in supporting users of whom they are unaware, and who may be using the information in very different ways from those originally envisioned? The web supports the collaborative activities of small work groups, but it also supports those of well-defined scholarly sub-disciplines, and those of the much less defined community at large. To study and support collaborative activities on the web, we need a methodological framework that identifies the major distinctions between 'work groups' of widely differing sizes and structures, and between the various roles that originators, retrievers, and intermediaries can play.

This article identifies one of the major problems of collaboration on the web as that of maintaining *awareness* between remote research partners when changes occur in one location that affect activities in another. Such *chronological awareness* is an important issue for supporting task-oriented collaborative projects of research groups or organizations. At the other end of spectrum, the issue of locating where specific information resource is on the web, i.e., *resource awareness* has become important for supporting the research community at large. In the subsequent sections, awareness issues in collaborative group are discussed, an awareness maintenance framework is developed which identifies the various dimensions for evaluating and designing awareness maintenance systems. CHRONO, a tool developed to support awareness of change, is described, and the awareness maintenance framework is used to analyze and classify this and other chronological awareness systems on the web.

COGNITIVE ARTIFACTS AND SITUATIONAL AWARENESS

The use of social interaction and cognitive artifacts as means to enhance human abilities has been analyzed by Norman [Norman91]. A cognitive artifact is defined as "*an artificial device designed to maintain, display, or operate upon information in order to serve a representational function.*" The power of a cognitive artifact comes from its function as a representational device changing the nature of a task to match human abilities and hence to give users *greater capabilities based on those abilities*. When the informational and processing structures of the artifact are combined with those of the users, the result is to expand and enhance the cognitive capabilities of the total system of human, artifact, and task.

CSCW (computer-supported collaborative work) research focuses on assisting people to work collaboratively as a cohesive team and providing them with a sense of common purposes (e.g., completion of the group task). For example, Landow's *In Memoriam* project [Landow90] utilizes hypertext's freedom of navigation and linking ability to alleviate the effects of physical separation and the univocal voice of textual conversation. In so doing it creates a new awareness of the processes of collaborative learning and collaborative work for group members in literary studies. Olson and Atkins' NSF EXPRES Project [Olson90]uses intelligent, multimedia email to facilitate cooperation within scientific and engineering community by increasing researchers' awareness about each other's work.

Norman notes that a critical requirement in shared tasks is maintaining *situational awareness* [Norman91] by keeping everyone adequately informed. In an environment where each member has a well defined role, the need to have face-to-face communication in order to perform a cooperative task becomes less necessary if mechanisms for situational awareness have been well established between members. For example, Hutchins notes that the navigation of a large ship requires effective coordination of various people with differing roles [Hutchins90]. Many key members of the navigation team are geographically separated and communicate with each other by a common telephone circuit. Those arrangements provide opportunities for the navigation crews to observe each other's work, contributing to partial redundancy in their joint knowledge. They also support maintenance of the group over time to provide fault-tolerance if some group members fail to perform their roles.

Thus, one of the important criteria for achieving group cohesiveness is the situational awareness of what other group members are doing. Together the functional specificity of the crews and the cognitive artifacts that facilitate situational awareness (e.g., the single telephone circuit, the pilot house with high visibility among navigational team members) can create an effective collaborative system. As already noted, for the web these awareness issues are made more complex by the diffuseness and lack of overall goal and task definitions for many of the 'work groups' involved.

FRAMEWORK FOR HUMAN FACTORS OF AWARENESS MAINTENANCE

This section examines various human factor issues relating to awareness of changes in a web-based working environment. The main focus here is on *chronological awareness* as a special case of situational awareness, that is, the awareness of *when* something (an event or an artifact) has changed.

Web pages, FTP archives, Listservers, and other common Internet infrastructures are the primary means for information dissemination on the web, and these infrastructures are being constantly updated to reflect members' current states of knowledge on their portions of collaborating tasks. In a dynamic environment where large amounts of information are created and updated frequently, the need to keep up with the most up-to-date and relevant information has become more important as the Internet community expands.

DIMENSIONS OF AWARENESS MAINTENANCE ARTIFACT

There are four main dimensions of design considerations for awareness maintenance artifacts for web users:

1. Locus of Responsibility: *Server-Side, Client-Side*, or *Centralized Dispatcher*
2. Level of 'Work Group' Hierarchy: *Group*, *Organization*, or Community.
3. Method of Locating Changes: *Browsing* vs. *Targeting*
4. Complexity of User Interaction: *Simplicity* vs. *Customization*

The first dimension, the *locus of responsibility*, differentiates who is responsible for maintaining the record-keeping mechanisms for supporting awareness maintenance.

- A *Server-side* approach ensures that only the users who are current visiting the web site would need to know what information has been changed. Hence, it reduces network traffic by avoiding needless broadcasting of chronological information to some users who might not be concerned with it. The main disadvantage of this strategy is that in order to know whether or not any particular page has been changed, a user would need to check out the specific web site periodically.

- A *Client-side* approach periodically monitors specific pages at various web sites and report whether or not they have been changed recently to ensure the user would be aware of any changes. The main disadvantage is that the users need to remember to run such a system, or it must be set up to run periodically, in turn causing it to consume higher network bandwidth.

- A *Centralized-dispatcher* approach put the monitoring responsibility at some specific central registry locations that automatically monitor the registered pages for the users. Its main disadvantage is the high network traffic involved in such a centralized broadcasting scheme.

The first dimension can also be considered from a information resource provider/user perspective. Various information systems, such as the web, use the client-server model to partition the computational division of labor. Similarly the *locus of responsibility* of awareness maintenance at every level can be divided into *originators* (i.e., providers) of information, *retrievers* of information resource, and *intermediaries* of information retrieval exchange. Therefore:

  - An *originator* is a source/server of information dissemination.
  - A *retriever* is an user/client of information resource.
  - An *intermediary* is a meta-information resource mediating between originators and retrievers.

The second dimension, the *level of 'work group' hierarchy*, reflects the need for maintaining mutual awareness among members exist in various collaborative arrangements. There are three main levels of awareness arrangements which constitute the awareness maintenance hierarchy.

- The *Group* level usually consists of closely coordinated members in a relatively small or medium size project. There is usually a need to be aware of short-term changes in the data being managed by different group members as part of their task activities, and the nature of this need is relatively well-defined.
- The *Organization* level frequently involves loosely coordination of different group projects within

an organizational structure such as a scholarly sub-discipline. There is usually a need to be aware of significant changes in the knowledge available in the organization, and satisfying this need involves mechanisms ranging from well-defined channels, such as electronic journals or reprint archives, to organizational/cultural/ethical norms on how information is to be disseminated.
- The *Community* level often involves mechanisms for providing its members at large some form of *resource awareness*, for example in locating where specific information resource is on the web.

Viewed from Miller's *living systems theory* [Miller78] perspective, *situational awareness* is essential at every level above the individual level in the system hierarchy and that originators and retrievers of information are situated at the opposite ends of the *channel* and *net* information subsystems [Chen96]. The various awareness maintenance mechanisms (such as the chronological awareness support systems for groups) serve as (or provide the functionalities of) the *timer*, *associator*, and *memory* subsystems in various system levels.

The third dimension, the *method of locating changes*, involves two different ways of locating documents that have been changed: browsing and targeting.

- A *Browsing* approach facilitates the chronological browsing characteristic: visitors of a site may find relevant information via browsing the concurrently created/modified web pages, because closely related documents are sometimes created (or modified) around the same time. This browsing approach allows the *at glance* attribute for accidental discovery of relevant information without prior awareness of their existence.

- A *Targeting* approach focuses on specific pages or information that users have previously specified. Therefore this method of locating changes is more direct and efficient. However the main disadvantage of such approach is that the users cannot be made aware of any new information which have been created recently. They are limited only to changes made to prior knowledge.

Finally, the fourth dimension, the *complexity of user interaction*, denotes system usability in terms of simplicity vs. customization.

- A *Simplicity* approach tries to make its user interface simple and familiar to web users. The goal is gearing toward ease of use and a shallow learning curve. On the web this is typified by systems that present awareness information through the generation of familiar web documents.

- A *Customization* approach, in contrast, tries to allow elaborate customization of features, but it also demands more efforts by its users to learn and utilize its functionalities. On the web this is typified by systems that use separate tools with their own user interfaces to support awareness.

CHRONOLOGICAL AWARENESS SUPPORT SYSTEMS

This section presents a concrete example of a class of awareness support systems specially designed for supporting *chronological awareness* maintenance on the web. One such system is described in detail, followed by brief overviews of three other systems.

By definition, a *chronological awareness support system* for the web provides each individual an appropriate awareness of relevant activities of other individuals. Hence it allows team members to synchronize their activities in a more coherent way by keeping them informed and aware of any changes made to each other's web pages or other information resource that might be relevant to their current tasks. The information systems that support such awareness correspond closely to the timer critical subsystems in a living system [Miller78, Chen96].

THE CHRONO CHRONOLOGICAL AWARENESS SYSTEM

CHRONO is an HTTPD server-side system which generates chronological listings of web pages that have been changed recently at specific sites. It provides a basic awareness-support that lets visitors of a web site (e.g., members of a group, or other Netsurfers) see which web pages have been modified since their last visit. Currently, the CHRONO system is implemented on a UNIX platform. CHRONO presents visitors with an HTML document that lists the titles of web pages at the site in reverse chronological order. This chronological listing also functions as a collection of hyperlinks to the listed web pages.

CHRONO USER INTERFACE AND ITS FUNCTIONALITY

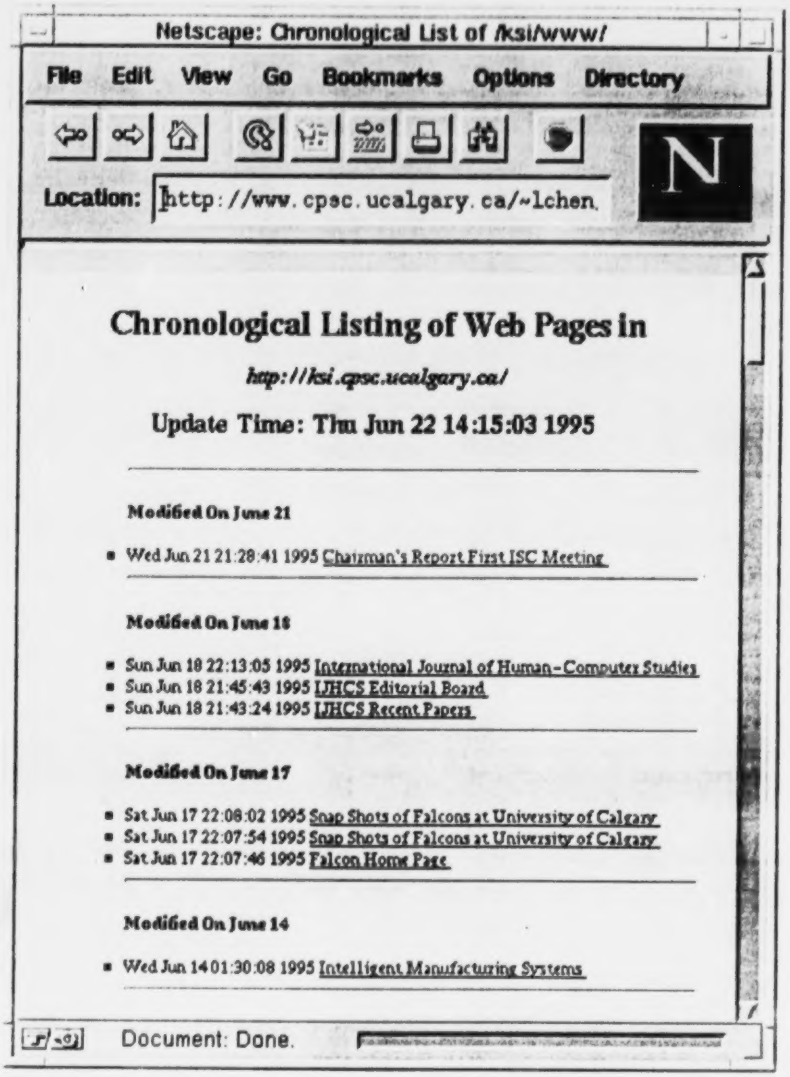
The user interface of the CHRONO system is straightforward and intuitive for web users. It looks like an automated *what's new* page to the users. From the list, the visitors are able to tell at a glance what documents have been modified or created recently. They can also scroll down the list to check those older documents at the site. Because the titles of the listed pages also act as hyperlinks to the actual web pages, visitors can simply click and jump to the relevant pages of interests.

Chronological hyperlinks presented in the CHRONO listings provide the visitors the means to access the newly modified or created pages. This *time-line* (or history) dimensionality complements the functionality of the associative memory characteristic found in typical hyperlinks which join related information.

The *time-line* dimension allows frequent visitors of a web site an immediate awareness on what have been changed since their latest visit. The changes may reflect some web pages in which they have been previously interested or they may show some pages that the visitors have never seen before but now appeal to them. Hence this *chronological* browsing characteristic is analog to *spatial* (subject-category) browsing characteristic that library patrons have often experienced when looking for books on open book-shelves (i.e., accidentally finding other relevant books near the books that they were looking for originally). Unlike a manually updated *What's New* page in which the users have to rely on timely updates made by a webmaster (or by the document authors), CHRONO provides the time-line dimension to the users automatically, in a reliable and periodic fashion.

CHRONO CURRENT SYSTEM USAGE

The CHRONO system has been running at the Department of Computer Science at the University of Calgary since March, 1995. It currently services seven web locations: two research units and five individuals. People associated with the two research units: Knowledge Science Institute and GroupLab periodically have utilized the system to check on new developments of each other (both within group and between groups).



Location: http://www.cpsc.ucalgary.ca/~lchen.

Chronological Listing of Web Pages in

http://ksi.cpsc.ucalgary.ca/

Update Time: Thu Jun 22 14:15:03 1995

Modified On June 21

- Wed Jun 21 21:28:41 1995 Chairman's Report First ISC Meeting

Modified On June 18

- Sun Jun 18 22:13:05 1995 International Journal of Human-Computer Studies
- Sun Jun 18 21:45:43 1995 IJHCS Editorial Board
- Sun Jun 18 21:43:24 1995 IJHCS Recent Papers

Modified On June 17

- Sat Jun 17 22:08:02 1995 Snap Shots of Falcons at University of Calgary
- Sat Jun 17 22:07:54 1995 Snap Shots of Falcons at University of Calgary
- Sat Jun 17 22:07:46 1995 Falcon Home Page

Modified On June 14

- Wed Jun 14 01:30:08 1995 Intelligent Manufacturing Systems

Figure 1 Screen Snapshot of CHRONO

From a preliminary examination of the HTTPD access_log of the CHRONO system site and from talking with individual group members, we have found that the chronological listings of the five personal sites have offered other group members more focused chronological awareness about these five individuals' working patterns. Occasionally, some people had discovered new projects the targeted individuals were working on that they were not previously aware. This positive usage experience of personal chronological listings suggests to us that there is a need to further examine the effectiveness of providing different sub-groupings of chronological listings to group members.

RELATED CHRONOLOGICAL AWARENESS SYSTEMS

There are three other systems that provide support for chronological awareness based on different designs and implementations. They are briefly examined in this subsection, then followed by a comparative evaluation.

WEBWATCH

BEST COPY AVAILABLE

WebWatch is a client-side chronological awareness system for keeping track of changes in selected web documents. Given an HTML document referencing URLs on the web, it produces a filtered list, containing only those URLs that have been modified since a given time. The criteria used for filtering can be given as a global setting that applies to all URLs, or can be derived automatically, using the time of user's last visit to the document, as recorded by the web browser in the user's local HTML (e.g., bookmark) file. In contrast with the simple time-line listing strategy used in CHRONO, WebWatch stores its arguments in a parameter file. Once the users have customized the program to their needs, using its graphical front-end, they can have it run periodically in unattended mode.

KATIPO

Katipo is another client-side chronological awareness system built for Macintosh that shares many similar concepts with WebWatch. It reads through the Global History file maintained by some web browsers checking for documents that have changed since the last time a user viewed them. The basic difference between it and WebWatch is that it uses the Global History file as its reference for checking URLs, whereas WebWatch uses the Bookmark file.

URL-MINDER

URL-minder is a centralized system that keeps track of resources on the Net and sends registered users e-mail whenever their personally registered resources change. Users can have URL-minder keeps track of any web resource accessible via HTTP. It keeps track of one web page, image file, or other Internet resource at a time. It tracks the actual HTML markup, binary contents, or ASCII contents of the URL users have submitted. If an HTML page includes a GIF or JPEG graphic, the URL-minder will inform them via E-mails when the reference to the graphic changes. The URL-minder currently checks on users' registered URL's at least once per week, and will inform users if it fails to retrieve their registered URL after trying twice.

SYSTEMS EVALUATION WITHIN THE METHODOLOGICAL FRAMEWORK

This section presents a comparative evaluation of the four chronological awareness support systems discussed earlier: CHRONO, WebWatch, Katipo, and URL-minder. Each system has its unique approaches for achieving chronological awareness for web users and complement each other along four main dimensions.

The first dimension, the *locus of responsibility*, differentiates who is responsible for maintaining the record-keeping mechanisms for chronological awareness. For example, CHRONO is a server-side system in which chronological listings are being updated and kept at the web server-side. Thus, CHRONO can be thought of as offering "*chronological awareness on demand*". WebWatch and Katipo, however, put the responsibility of maintaining chronological awareness on the client-side. Both client side systems periodically monitor specific pages at various web sites and report whether or not they have been changed recently. Finally, URL-minder requires its users to register at a centralized site.

The second dimension, the *level of 'work group' hierarchy,* signifies that all of the current chronological awareness support systems are mainly focused at the group level (and to some extend at the organizational level). Collectively, they are supporting chronological awareness of information resource typified by closely-coupled collaboration at the group level.

The third dimension, the *method of locating changes*, involves two different ways of locating documents that have been changed: browsing and targeting. CHRONO uses the browsing approach in order to facilitate the chronological browsing characteristic. Conversely, WebWatch, Katipo, and URL-minder employ a targeting approach in which they are targeted on specific pages or information that users have previously specified.

Finally, the fourth dimension, the *complexity of user interaction*, denotes system usability in terms of simplicity vs. customization. CHRONO and URL-minder are in the simplicity category; their user interfaces are simple and familiar to web users (i.e., scrolling list of hyperlinks and fill-in form of URLs and e-mail address). They are geared toward ease of use and a shallow learning curve. Both systems, however, have no capability for individual customization. In contrast WebWatch and Katipo allow elaborate customization of features, but they also demand more efforts by the users to learn and utilize their functionalities.

Therefore, each chronological awareness support system examined so far have various degrees of advantages and disadvantages along the four dimensions. CHRONO has the advantages of: (i) simplicity of user interface; (ii) supporting accidental discovery via its browsing characteristic; and (iii) server-side chronological awareness information on demand. It is nicely complimented by WebWatch and Katipo for their strength in the efficiency of targeting approach and customization capabilities. And finally URL-minder offers another unique service: it uses e-mail as its notification channel. This approach is useful for users who use their e-mail systems more frequently than web browsers. Together as a whole, these chronological awareness support systems have covered a wide range of approaches in respect to four major dimensions of chronological awareness support.

CLASSIFICATION OF EXISTING WEB SERVICES

The methodological framework developed in this article can be used to study a wide range of general awareness systems for the web. There are various awareness maintenance artifacts on the Internet that address different system levels. The following analysis examines and categorizes them along the two major dimensions: *level of 'work group' hierarchy and locus of responsibility*. Currently, *intermediary* mechanisms are still in a nascent stage of evolution (e.g. Universal Resource Agents), hence the present taxonomy does not yet reflect the intermediary locus of responsibility (Table 1).

At the *group* level, originators of the information resource can organize and implement work flow models of the group activities, use server-side chronological awareness support systems such as CHRONO, or/and send e-mail notification to users of information. Retrievers of the group level information resource can use client-side chronological awareness tools such as WebWatch and Katipo or register in centralized dispatcher service like URL-minder. Alternatively, they can send e-mail to inquire to information originators to see if any new things have come up.

Table 1 Taxonomy of Mechanisms Supporting Awareness Maintenance

	Level of 'Work Group' Hierarchy		
Locus of Responsibility	**Group**	**Organization**	**Community**
Originator	Work flow Model **CHRONO** E-mail to Notify	Broadcast to Listserver Register in **What's New** Announce in MOO	Register in **Yahoo** Initialize **Web Crawler** Establish HTML Links
Retriever	**WebWatch/Katipo** **URL-minder** E-mail to Inquire	Participate in **HyperNews** Participate in MOO Follow HTML links	Browse **Yahoo** **LYCOS** Search Read/Post to USENET **Web Crawler** Search

At the *organization* level, originators of the organizational resource can broadcast to concerned individuals, groups, or organizations via specific Listservers, register in NCSA's *What's New* service, or announce in organization-maintained MOO or MUD. They can also establish what's new HTML links in organization news while retrievers can participate in HyperNews or MOO, and follow the new HTML links in organizational web pages.

At the *community* level, originators can register the information resource in hierarchical subject services like Yahoo, or initialize their pages in searching and navigational services like Web Crawler while retrievers can browse Yahoo, search LYCOS or Web Crawler, or read and post to the USENET newsgroups.

One interesting observation can be made about the level of awareness in relation to the level of coordination. As the level of awareness moves from the group level to the community level, the need for closely-coupled coordination decreases among members. What happens in practice is the awareness maintenance becomes asymmetrical, rather than mutual, at the higher system level. However major awareness requirement continues. Resource providers may not need to be aware of who their users are, but the users' activities may be critically dependent on the status of the resources.

DISCUSSION AND CONCLUSION

The discussions in the previous sections have utilized the concept of *system levels* in *living systems theory* [Miller78, Chen96] and described an awareness maintenance framework for the web that focuses on the dynamic interactions between individuals, groups, organizations and communities. Within this framework, individuals constitute the fundamental level of analysis in the overall system. Self awareness provides a sense of identity, purpose, and consciousness to each individual. When perceiving all levels collectively, the notion of self awareness can be extended to different hierarchical partitions in the system for the purpose of analyzing the inter-relationship between the overall system and its parts in the *collective stance* [Gaines94].

Within in this collective stance perspective, individuals, groups, organizations, and communities together create a multi-leveled *collective intelligence* [Smith94] that becomes a pool of human knowledge for individual collaborators in remote sites to share their ideas and all aspects of collaboration. In order for the collective intelligence to function in a cohesive and integrative manner, coordination at various system levels must be maintained. However the means for maintaining a system-wide cohesion and integrity for the web are still emerging and evolving.

One practical issue inspired from the locus of responsibility for awareness maintenance at the community level is that of awareness of individual member's profile of specialties as a information resource. For example in terms of scientific communities, one common question among many scientists/scholars is: who and how should one contact when specific research questions, needs, or opportunities have arisen? The challenge is how can we create an environment inductive for informal social networking among community members on the web? What is the proper balance among originators, retrievers, and intermediary mechanisms for such an information resource?

Also in term of chronological awareness, there is a need to examine further the effectiveness of providing different sub-groupings of chronological listings to group members, such as chronological listings for particular sub-directories of a web site, for particular projects of a group, or for chronological access patterns to particular documents which may prove to be useful on different occasions and for varying user needs. In addition, the issues of how to establish mutual awareness among individuals, groups, organizations need to be explored further.

REFERENCES

[Berners-Lee94] Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F. & Secret, A. (1994). The World-Wide Web. Communications *of the ACM*,August, Vol. 37, No. 8, pp. 76-83.

[Chen96] Chen, L. L.-J. & Gaines, B. R. (1996). A CyberOrganism Model for Awareness in Collaborative Communities on the Internet. *International Journal of Intelligent Systems*, to appear.

[Gaines94] Gaines, B. R. (1994). The Collective Stance in Modeling Expertise in Individuals and Organizations. *International Journal of Expert Systems.* 7(1), pp. 22-51.

[Galegher90] Galegher, J., Kraut, R. E. & Egido, C., (ed.) (1990). *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work.* Erlbaum, Hillsdale, New Jersey.

[Hutchins90] Hutchins, E. (1990) The Technology of Team Navigation. In [Galegher90], pp. 407-428.

[Landow90] Landow, G. P. (1990) Hypertext and Collaborative Work: the example of Intermedia. In [Galegher90], pp. 407-428.

[Miller78] Miller, J. G. (1978) *Living Systems*. McGraw Hill, New York, NY.

[Norman91] Norman, D. A. (1991) Cognitive Artifacts. Carroll, J. M., (ed.) *Designing Interaction: Psychology at the Human-Computer Interface*. pp. 17-38. Cambridge University Press, Cambridge, UK.

[Olson90] Olson, G. M. & Atkins, D. E. (1990) Supporting Collaboration With Advanced Multimedia Electronic Mail: the NSF EXPRES project. In [Galegher90], 429-451.

[Schatz91] Schatz, B. R. (1991). Building an Electronic Scientific Community. *Proceedings of 24th HICSS-- 1991,* Vol. 3.

[Smith94] Smith, J. B. (1994). *Collective Intelligence in Computer-Based* Collaboration. Lawrence Erlbaum Associates, Hillsdale, NJ.

This paper was presented at the WebNet 1996 Conference, October 16-19, 1996, San Francisco, CA; Association for the Advancement of Computing in Education (AACE).

---

guest book | email

Last update: 1996 07 18 by Lee Chen &ltlchen@cpsc.ucalgary.ca>

Innovative Resources for Education and Public Information: Electronic Services, Data and Information from NASA's Hubble Space Telescope and Other NASA Missions

Dr. Carol A. Christian
Director, Office of Public Outreach
Hubble Space Telescope Mission
Space Telescope Science Institute
Baltimore, MD, USA
carolc@stsci.edu

Abstract: The Space Telescope Science Institute, which supports the operation of the Hubble Space Telescope, is actively investigating and supporting innovative and experimental methods for improving science and math education content. The educational resources on the Web are derived from the latest data, scientific results and the advances in the supporting technology of the Hubble Mission. The resources and services offered are created through strategic partnerships between scientists, technical staff, educators, informal science institutions and other key organizations. The inspirational nature of the astronomical data lends itself well to multi-media applications and methods using new Internet technologies that cultivate exploration and discovery. The experiments conducted in bringing the scientific content, the human expertise and the technology together for the benefit of the public and the educational community are illuminating with regard to the barriers and cultural differences between the various participants. Several examples of Internet based programs and activities are illustrative.

Introduction

Research scientists are increasingly aware that the public perception of the scientific process and the development of technology is frustratingly incomplete. Yet most people would admit that the functioning of our society depends upon developing an educational system that produces a literate workforce. Included in the skills that are needed in society are thinking, planning, and reasoning abilities along with agile navigation through information sources and manipulation of computer resources. Electronic communications and information technologies offer a new educational opportunities for individuals to experience and become adept at inquiry-based learning and discovery, both through self-guided means and through collaboration and coaching by educators, scientists and technically-oriented persons. Some aspect of such skills, from casual browsing of electronic materials to maneuvering through time critical electronic transactions, is relevant in every arena and industry.

Why Astronomy?

An area of scientific research that is fascinating to most people at intellectual as well as emotional levels is astronomy. Fortunately, observations of the cosmos coupled with the intricate analysis techniques which reveal the underlying physical nature of the objects studied lend themselves well to appealing multi-media applications. Contemplation of the technological advancements and instrumentation that astronomers deploy offers unlimited opportunities for creating enticing instructional tools. By examining the science data, information and expertise which is rapidly made available through on-line services to the scientific research community, one might expect that the underlying techniques and methods can be re-shaped and adapted to engage a broader audience such as the news media, the general public, businesses, educators, students and science and technology enthusiasts in the use of emerging technologies.

Astronomy and Space Science Relevance to Web Based Applications

Astronomical data is hard won in its highly competitive field - scientists acquire data through experiments conducted with high demand ground-based observatories, unique instruments on orbiting satellites and scientific probes launched into deep space. Due to the precious and competitive nature of the data acquired, and also the fact that each observational facility offers scientists access to a unique window into the physical universe, much astronomical material exists in on-line living archives offering access to information bases through relatively "standard" scientific formats or easily translatable formats. In addition, astronomy as a discipline has addressed the challenge of globally distributed information systems for nearly a decade. Accordingly, scientists have developed distributed computing infrastructure, search engines, location services, security mechanisms and conversion services to allow access through client-server and peer-to-peer technology to data which is distributed over a large number of host sites. The systems in use today have necessarily (by popular demand and intense fiscal pressure) migrated from the earlier proprietary systems to popular public access methods. Remarkably enough, a significant fraction of the astrophysics scientific community has agreed upon implementing Web based interfaces to the underlying, heterogeneous databases and archives scattered across the globe.

At the Space Telescope Science Institute, as for several other NASA missions, we strive to offer what were traditionally scientific resources to a wider audience who posses a large range in technical and scientific expertise. The obvious goal is to provide substance for educational curricula, but also to provide access to NASA resources for industry, the media, informal science centers (museums), and the general, international public, to name a few. The resources offered clearly can ignite interest in the particular science and discovery, but the data and information also can be woven together into imaginative tapestries to demonstrate emerging technologies. Specifically, astronomical content can be used for a plethora of educational purposes, for example to teach effective research techniques, mathematical concepts, scientific processes, historical perspectives, writing skills, appreciation of cultural differences and viewpoints. Beyond that, astronomical content can be used creatively to demonstrate technological methods: search engine functions, documentation techniques, graphical interfaces, interactive services, real-time collaborative services, security systems, database functions, distributed computing environments, etc. that can be mapped, by analogy, to other disciplines, especially those with less engaging content or information that cannot be exposed in a public forum.

The migration of resources and data from the arena of discipline experts to a more general clientele creates many challenges. In a distributed system created by and for the domain expert (e.g., scientists) the information to be accessed must be made much more locatable and understandable for general users with heterogeneous backgrounds. Material is useful in education if it can be located quickly through simple browse products (text, thumbnails, short clips) cleverly crafted as entry points to the more complete tiered structure of resources. Once found, resources must clearly be relevant to the user's purpose. A greater challenge has been to construct the browse products, demonstration materials and the core content itself in an engaging presentation which cultivates *new* uses, value added products, and further exploration into the Web. While the domain expert may be fairly unhappy with cumbersome interfaces, in fact, *experts often are strongly motivated to use awkward access methods if the content is pertinent to their needs*. However, domain experts, similar to most other users, are typically rebellious against using multiple interfaces or several disparate and complex software constructs to obtain necessary information.

Research Tools and Data - Migration

Motivated to broaden usage of science data to non-specialist users, we hoped that the relatively mature infrastructure ("middleware") built for providing distributed access to NASA data, analysis software, and ancillary information for scientists (e.g., the Astrophysics Data System [ADS] http://wwwads.harvard.edu/ or the Hubble Space Telescope Archive http://www.stsci.edu/) could be migrated transparently to the wider audience. However, *this was not the case*. The initial intent was to provide a simple Web based wrapper to the already existing scientific interfaces because it was speculated that fairly light weight Web services could be tailored easily for different clientele. In practice, however, we have found this is not trivial. First, access into the full set of scientific materials is hopelessly complex for the non-specialist, even when driven through a simple interface. A deeper technical difficulty arises from initial lack of "state" in Web services, so that it was not possible to retain many of the sophisticated distributed computing attributes of the original astronomical data systems in use. Therefore subsets of NASA data systems are available through the Web, but unfortunately now without the backbone of middleware which once allowed scientists or anyone else to have "the universe on their desktop". An example of one of the sub-systems may be found through the "catalog service" at the ADS URL given above. Currently, several NASA projects are conducting experiments to grapple with these problems where one of the key goals is to provide ready access to resources without enormous and costly expenditures on infrastructure, interfaces and other re-design which would tax the severely diminishing resources in the missions.

Space Telescope Science Institute (STScI) Specific Application

The Office of Public Outreach (OPO) at STScI is responsible for disseminating scientific results, telescope and satellite information and other materials to all sectors of the public including educators. The other divisions at STScI concentrate on interaction with the science community and support the planning, scheduling, data acquisition, analysis and publication of results within the scientific research domain. To provide information in diverse ways to a broad audience, we are convinced that *direct involvement of the user in the design of services is critical*. In our experience with information technology industry, there is much lip-service given to such a strategy, but in fact, "representative users" are often allowed to be little more than temporary critics. User involvement *in the design* means that content relevant to a specific constituency is derived collaboratively through compromises and balance. *Knowledge of this approach frequently tempts the software engineer or scientist, "who knows better", to disengage from the process before it starts*. Therefore, managing collaborative projects is a challenging proposition and must include appropriate brokers to engage and mesh the expertise of the various participants productively. *Also, the research community must work conscientiously to demonstrate access to information infrastructure is actually useful, rather than acclaiming it so*. Therefore, OPO and other NASA supported programs have teamed educators and science museum personnel with scientists to provide engaging content that easily demonstrates the relevance of HST and other flight program materials to the public. Teachers, in particular, can then work cooperatively to appeal to school, local and state officials to obtain network access within their classrooms and computer laboratories. Another significant driver in working to develop an Internet savvy clientele is that dissemination of hard copy and other collateral material is expensive for any organization, especially since much of this material is originally generated in electronic form! It is far more effective to disseminate appropriate materials electronically, allowing users to pick and choose what materials they wish to obtain rather than providing large packets of collateral only a fraction of which will be used.

Strategic Partnerships and Teaming Are Critical: Cutting edge data and information from scientific investigations *should* naturally cascade through university graduate and undergraduate courseware to the K-12 community. At STScI, a research institution, close association with universities and community colleges starts to address this need. OPO/STScI support also is directed to the news media, to journalists providing public information, to science writers and to the science museum community. The initial hurdle is to educate those constituencies in the use of electronic information technology, and further to integrate news material into thematic packages which also contain materials appropriate to informal science and educational settings. *Organizing scientific data and resources along thematic lines is not trival* - archives and other information are often configured for particular flight missions, research objectives, or a specific scientific or technical clientele. Accordingly, the wide variety of types of users that OPO/STScI must support has offered the opportunity to experiment with a number of approaches for providing services. One of the meaningful (if not self evident) results of working in partnership with the users in the initial stages of specification of Web resources, is that we affirm that *access to human expertise, personal experience, face-to-face interactions of developers, information brokers, and users, and initial period of building of rapport is essential, no matter how engaging or astonishing the technology to be used is. The building of trust and credibility is important if expensive multi-media resources are to be any more than of fleeting interest to users.*

Presentation - Effect on Understanding: Another arena of investigation is to understand how the use of interactive technology (such as Java applets, Shockwave, etc.) conveys information, and what information is added, lost or altered when such technology is not used. This is often relevant for the competitive world of the news media and public information, where the "angle", that is, the specific presentation of content, can impact the effectiveness of a news story or journal article. In education, such issues are even more relevant. It is not always evident what presentation is the most effective. For

example, in the presentation of the new map of the surface of Pluto - assembled through analysis and modeling of a series of "snapshot" high resolution HST images - the whirling planetary globe is an attention grabber, and clearly by using it, the TV journalist has the edge over a newspaper writer. In education, the snapshot images, the Mercador projection, and the globe all represent the same information, but also convey information on how a mosaic of images becomes a map, how globes are projected into flat surfaces, how scientific modeling relates to actual data, and finally how HST is used to study rotating objects. These concepts can be described at length textual form, but rarely developing quick insight for the viewer/reader.
Innovative Processes: Another issue which is not well understood yet is: how important is novelty for multi-media in education? Certainly, engaging presentation is a time honored tactic to grab the learner's attention, but educators can ill afford to compete with the game industry to retain student interest. What are important are not only services that are innovative in their own right, but rather what *imaginative education processes* are built on interactive tools. For example, the recent "Live from Hubble" experiment involved a suite of on-line journals, chat sessions, interactive consultations with scientists, and rapid download of planetary images from HST, which were all inventive resources on their own. However, the combination intertwined into the overall program gave students and teachers a bond to STScI and to each other in a collaborative learning environment that was sustainable over several months. This would not have been achieved through use of any single tool, no matter how ingenious.

Other Issues

Clearly there is a plethora of educational and social issues surrounding Web use. One problematic area for educators, which directly impacts the work conducted at STScI is intellectual property rights and copyrights. While it is true that the rules of fair use usually state that employment of NASA materials for educational or non-profit purposes can be permitted or licensed, there remains a broad spectrum of interpretation of the law. The ambiguity is troublesome for educators and content providers who intend to exploit the Web for effective dissemination and access to resources, but have no intention to reap huge profits from the information accessed. Other organizations would be well advised to follow the lead of NASA missions in carrying an explicit statement that allows use and reproduction for education, in addition to their copyright notices. This is a great motivational factor for educators. OPO/STScI permits use of Web resources for education and research and pro-actively encourages widespread utilization of the dramatic imagery from HST (c.f. the copyright notice at the STScI web site: http://www.stsci.edu/).

OPO/STScI recognizes that our customers enjoy a large range in connectivity, technical expertise and desktop computer power. We have begun our own study regarding the resources which may need to be cached either at remote sites or downloaded frequently by users who need rapid recurrent access to some materials. For example, STScI serves a significant number of international users, so building mirror sites is a logical step, but some sites have capacity for only a limited amount of the high demand resources. Furthermore, we are studying which materials teachers prefer to download in advance of scheduled classroom activities so that classroom access is relatively rapid. Clearly, caching of information will be a relevant topic for some time to come in situations where distributed systems serve a full range of connectivity including global markets.

The experimentation conducted within the astronomical community in partnership with science museums, teachers, students, universities, community colleges and libraries is primarily for the benefit of teachers, students, and informal science learners. However, the programs offer a rich suite of lessons learned on how to approach different clientele in a cost effective way, which amazingly are repeated time and time again by other research institutions and the industry. Research institutions and university departments cannot expend precious research resources to create a plethora of types of packaging, wrappers, services, and other ancillary materials to reach the public. In the case of STScI, the public's thirst for materials quickly overwhelms the available support structure. Technical training and access to infrastructure are still problematic for the general user, and *it is not clear to the scientific research community how receptive the public will be to newly marketed products and services without sufficient training and education.* It is evident that as the ranks of the research community shrink as fiscal support dwindles, only small, focused education programs will be feasible.

A Few Samples of NASA Funded Web Based Programs and Resources

Program	*Description and Link*
Space Telescope Science Institute (STScI)	Public Information, News, Education Resources, Classroom Modules - http://www.stsci.edu/public.html
Science Information Infrastructure	Educational Resources, Lesson Plans, Classroom Activities. Partnership - Researchers, Science Museums, Industry, Educators - http://www.cea.berkeley.edu/Education/SII/
Remote Sensing Public Access Center (RSPAC)	Publicly Accessible Data and Resources, Links to NASA supported Programs in Space Science and Earth Science, Supports NASA Information Infrastructure and Technology Program (IITA) - http://www.rspac.ivv.nasa.gov/
Other IITA Programs	Aeronautics, Digital Library Initiatives, K-12 Programs - http://rsd.gsfc.nasa.gov/rsd/iita/IITAproj.html

Summary

Using astronomical research as a context, and drawing upon almost a decade of experience in distributed computing techniques, the astrophysics community is well placed to experiment with a variety of techniques, methods and technologies for Web use. Interaction with industry would be beneficial for both communities. OPO/STScI serves a wide variety of users (news media, scientists, general public, teachers, administrators, students, parents, etc.) who require access to resources at different levels of complexity, on different timescales and through several types of interfaces. The advantage of astronomical content is that most individuals find it motivational so that experiments addressing some key issues relevant to information technology and Web use in a variety of situations find welcome participants. The creation of resources can be accomplished successfully through strategic partnerships with knowledgeable brokers, educators, science museum personnel and others. In addition, innovative resources are useful only if they are interwoven into imaginative and engaging processes that have learning as the primary end product.

We continually re-learn and reinforce the idea that strategic partnerships that involve suitable information brokers result in the best migration of science and technology to the public. These brokers must not only be knowledgeable about domain specific content and possible new uses, but also must be sensitive to the "cultural" background of the participants. For example the scientific and technical fields are highly competitive, often introspective and rarely easy to interface with whereas the non-expert demands ease of use, a tiered structure of information and engaging materials. The NASA mission is expressly to foster research, experimentation, innovation and unusual or risky technology development. NASA funded programs cannot and should not be aimed at duplicating efforts initiated by other federal agencies or industry, and in particular can support but not drive systemic education reform.

Acknowledgments

This work has been in part by NASA contract NAS5-30180, grant NAG 5-2875, and contract NAS5-26555. Many individuals at several institutions have participated in the various programs contributing to the material represented herein and their contributions are gratefully acknowledged.

The Web as Process Tool and Product Environment for Group-Based Project Work in Higher Education

Betty Collis*
Toine Andernach**
Nico van Diepen**

*Faculty of Educational Science and Technology
collis@edte.utwente.nl

**Department of Computer Science
{andernac, nvdiepen}@cs.utwente.nl
University of Twente
Postbus 217,
7500 AE Enschede, The Netherlands

Abstract: We discuss problems confronting the use of group-based project work as an instructional strategy in higher education and describe two courses in which course-specific Web environments are used both as tools for group-process support and as the product environment of the project work. In particular we describe specific Web-embedded shared workspace, communication management and evaluation tools and their contribution to the management and educational value of group-based project work. We argue that both pedagogical and technical strategies are needed for efficient and effective support of group-based project work in higher education.

Introduction: Group-Based Project Work as an Educational Strategy

Group-based project work, or project-based education, is a form of instructional organization with a long history. It can take many forms, varying on key dimensions such as the degree of openness in the problem being addressed via the project activities, the choice of individual or group engagement as the mode for carrying out project activities, and the relationship of the project to the course in which it is situated. For our own work, we define project work in the university context as involving groups of students who must collaborate or at least co-operate in the execution of a multi-phase activity designed to embody the learning goals of a course and involving at least the stages of problem analysis, decision making about approach choices to problem-related actions, design and production of problem-related output, and iterative evaluation of output in relation to the original problem and the goals of the course.

Key Variables and Research Questions Relating to Group-Based Project Work

Within these phases of project work, many different variables influence the educational value of the experience and certain key problems repeatedly occur. Project work involves process as well as product; process must be managed both internally and externally, with attention to efficiency and group memory; group-interaction variables affect both process and product; and the product of project work must fit the problem or needs that stimulated the project activity. All of these are complex tasks, particularly for the instructor also involved in the execution of the project activity as part of a broader course experience. When some or all of project-work activity takes place among participants distributed over time or distance, the challenges involved in making project work a productive and manageable instructional strategy increase [Collis & Smith 1996]. A general question of broad interest is thus:

Which pedagogical and management strategies, combined with which Web-integrated functionalities, improve the efficiency and educational effectiveness of group-based project work for university students in design-oriented technical courses where members of the courses include students from different *countries and educational institutions?*

Two courses in which we are involved and which make project work a key instructional strategy are serving as cases for us in our on-going examination of this question, both to improve our own teaching and in relation to a joint research focus on "tele-learning" involving both of our faculties at our university. In this paper, we describe the 1995 versions of these two courses, one for educational technology students and the other for business information technology students; indicate some key problems addressed by (ourselves as) the instructors of these courses; describe the Web-based tools and environments designed and used by the instructors within the courses in support of the project work; and reflect critically on future refinements in project-work deployment in our courses

Description of the Courses

The course *On-Line Learning* is an elective course for senior students in the Faculty of Educational Science and Technology at the University of Twente in The Netherlands. Students electing the *On-Line Learning* course generally are specializing in the design and development of electronic instrumentation for learning support. In the 1995 version of the course, there were 32 students including five exchange students from Spain and Germany. Relating directly to the objectives for the course is a collaborative learning activity, around which the project work is based. In 1995 this collaborative learning activity was the design and development of a hyperlinked reader making effective use of the functionalities available in the World-Wide Web environment.

The Jigsaw Methodology for collaborative work was employed, by which each member of a group has a specific task which must be contributed in an appropriate fit to the overall "jigsaw" of the project [see, for example, McManus & Aiken 1995]. In the groups of three, all group members had to agree on the structure and global design decisions for their component of the communal Web site, but after this it generally proceeded that one student took the lead for the actual technical development of the site, one for the content, and one for design decisions relating to layout and navigation. The instructor managed the project work through a mixture of computer conferencing using the FirstClass computer conferencing environment, e-mail, and distribution-list communication. Over 500 messages were logged during the course. It was required that all group-produced components be integrated into the "new textbook" as a total learning resource and that the entire environment be completed and presented (to the class and to the "world" via its movement to a WWW server for Internet access) during the final face-to-face session of the course. The product of this project work can be seen within the 1995 course site, at

URL http:// www.to.utwente.nl/ism/online95/campus/campus.html

Applications of Information Technology (*AIT*) is a compulsory course for first-year students in Business Information Technology (BIT) at the University of Twente (the Netherlands). The Business Information Technology curriculum started in 1993 to bridge the gap between computer science and business administration. *AIT* makes this very explicit, by using information technology to study applications of information technology. *AIT* provides an introduction to the full spectrum of applications of information technology, allowing students to zoom in on specific applications. The course emphasizes the role of the technology in organizations rather than the technology itself. Thus, the course serves to motivate the students for what is coming in the remaining curriculum and makes sure that they can relate theory to practice. A special feature of the course is the experience of working collaboratively on the major course project with students taking a similar course at the Kuopio Vocational Educational Centre in Finland via the Internet.

Students work in project groups each of which consists of about six Dutch and three Finnish students. Each group has its own Web page that gives an overview of the group's work. Efforts of the group are to yield a collaboratively produced report about a selected topic, in the form of a set of Web pages in which individual contributions are integrated. The individual contribution of each member of a project group consists of a page in which that member presents himself, and a number of pages in which the acquired knowledge is presented in the form of an essay. The course takes nine weeks each consisting of two sessions of four hours each. A group arranges its own work, making sure that the next week's commitments are kept. Access to other equipment or access beyond hours is possible and may be necessary to achieve the desired goals. Tutorial sessions for the course stress the relation of the project to the content of the course textbook. Supervision contacts are maintained with each local subgroup, to check progress; to account for the commitments due; to stimulate improving the work; and to ensure that delivery of the final document is guaranteed. Self evaluation by the students is carried out electronically by means of an evaluation tool embedded in the course Web site.

After having successfully followed the course, students' knowledge of the role of information systems in organizations, failure and success factors of applications of information technology, information sources and Internet applications should have increased. Furthermore, students should be able to formulate an information request, to collect information, to use information technology for collaborative document processing and to co-operate effectively. International co-operation, taking initiatives in order to achieve goals, thinking critically and constructively and respecting the requirements from within an organization are other important aspects of the course.

Use of Web-based Tools and Environments for Process Support in the *AIT* Course

In the 1995 version of the *On-Line Learning* course described here, the Web was used as the platform for the products of collaborative work and well as the environment that was the focus of study of the collaborative work itself. However, the Web site for the course was intended more as the product and archive of the course rather than as a tool for supporting the process of project work during the course. Separate communication environments supporting various forms of communication and file transfer were the tools for this support. In the *AIT* course, the Web is used both as a management and work site and a product-related environment, although the product is realized more individually compared to the *Online Learning* course. Of particular process support in the *AIT* course are functionalities integrated within the course Web site relating to mail archive, to shared workspaces, and to evaluation tools. In the following, the functionalities of these features are described and their integration into the overall *AIT* course site explained. Thus we discuss these in some detail here.

A Web environment was created which was explicitly meant to be a starting point for students following the course. The same environment was available for both the Dutch and Finnish students to assure the same basic information. This Web environment served several purposes:

1. providing practical information of the course: planning, time schedules, goals and phases were available
2. providing local course information such as manuals, papers and an example of a Web environment created by one of the last year's *AIT* groups
3. supporting the information-search process (links to the local on-line library, Web search engines and electronic journals)
4. supporting both synchronous and asynchronous communication among both local and international group members, with other groups and with and among the tutors
5. providing feedback on communication by a group memory (mail archive) for every group
6. supporting collaborative document processing by integrating shared workspaces into the Web site
7. presenting the (intermediate) results and the end products (Home Pages)
8. supporting self evaluation by students through an embedded evaluation tool

The tools for purposes 5, 6, and 8 are further described.

***Mail Archives*:**

Mail archives are designed to support the structured storage and retrieval of group mail messages. As such they function as the shared memory of the communication process between the group members. Messages sent to the archive are stored and distributed to the members of the archive. In *AIT*, every project group (including its tutors) has its own archive. The archive's Web interface allows its members to reply to messages already present. The interface has different modes, for example, messages can be ordered by thread, by subject, by date, and by sender.

Shared Workspaces:

Shared workspaces are environments for collaborative document processing. The tool used in the *AIT* course is *BSCW, Basic Support for Co-operative Work*, developed by the GMD Institute (German National Research Centre for Computer Science) in Bonn, Germany. BSCW supports the storage of several different kinds of documents such as texts, pictures, hypertexts, audio and video. It is Web-based so using the environment is platform-independent. Every group has its own workspace. BSCW keeps track of the events in the workspace, such as adding, reading, and updating documents and of the agents involved in those events. Features such as version management and file locking are also available.

Evaluation Tool:

An evaluation tool has been developed to support evaluation of the course by all students. The tool is integrated in the Web environment of the course. It consists of three components. The first is a questionnaire with 41 five-point Likert scale items that must be filled out and electronically submitted. The second component processes the data after receipt; several statistical analyses are possible but in *AIT* basic measures like frequency and median are used and histograms are created. The third part of the tool consists of the HTML document used to graphically present the evaluation results. More specifically, histograms show the absolute frequency of each score per item (split per nationality, Dutch or Finnish) and the median scores of all items (also split per nationality). Shortly after submitting the form, the graphical representations of the results can be accessed via a hyperlink.

Key Strengths and Problems in the Project Work

Although the courses have been positively evaluated by their students and instructors [see Bos, Morgan, & Kikstra 1996], [van Diepen & Pouw 1995] and [Pouw, Terlouw, Joosten, & van Diepen 1995], they highlight on-going problems as well as satisfactions in the use of group-based project work as an educational strategy. We indicate some of the major satisfactions and some of the major problems.

In the *On-Line Learning* course, key satisfactions came from the close collaboration and interdependency both within and between the groups, the ease with which the local Web environment allowed each group to see the evolving work of the other groups and use this as the basis for coming to consensus about key design decisions for the overall product. The students came to call the Web site their construction area and their "virtual campus", and gave themselves their own "rooms" (their metaphor for their home pages) and "living areas" (their metaphor for areas under construction. A number of these metaphors are shown and discussed in the site itself, via the link "Our Environment" from the homepage). The sense of common ownership of a product by 33 students was linked to a positive learning dynamic in the course [Collis 1996].

Some problems however did occur with the project work in the *On-Line Learning* course, particularly with respect to the efficient and effective management of project work, such as organization of group resources, the extraction and organization of key materials for a "group memory" for each project group that could be shared not only by each set of group members but also by the instructor and members of other groups. The text-based computer conferencing environment, on a separate server and highly constrained in the type of file it could handle from outside of its own message environment, lost its functional benefit even for asynchronous communication as more and more project-work communication came to be related to HTML and graphics files and ideas from other WWW sites.

In the *AIT* course, key satisfactions and problems in the management of project work occurred in several ways. The main satisfactions concerning the management of the project were the students' insights into the possibilities and limitations of international group-based project work, the integrated Web environment and the automatic evaluation tool. The Web environment in which starting points for communication, information search, collaborative document processing, presentation and evaluation were integrated, appeared to be very useful.

Although international co-operation was more successful in 1995 than in 1994 some problems remained. Learning goals must be shared by all group members However, Finnish students regarded the course as less relevant than Dutch students. This seems to be related to the different status of the course in the respective curricula and to the different student populations in the institutes involved. Differences in cultural and social background, educational system, mother tongue and motivation may cause problems in reaching the same learning goals. Especially in time-critical situations, e-mail as an asynchronous medium did not fulfill the communication needs in *AIT*. Furthermore, we experienced that BSCW, the tool we introduced for stimulating collaboration, was not very frequently used for this purpose.

Combining the Strengths

In terms of more generic applications to the improvement of group-based project work in higher education, the two courses both suggest the use of an integrated electronic work area employing functionalities such as those available in the Web combined with an effective pedagogy. Table 1 highlights particular strengths of the courses in relation to persistent problems in the educational deployment of group-based project work (note that a comment given in relation to one course does not imply its absence in the other course, only its relative dominance).

Persistent Problem in Group-Based Project Work	**Contributions from the *On-Line Learning* course**	**Contributions from the *Applications of Information Technology* course**
Problems in maintaining course cohesion and momentum as students become immersed in their respective projects	Communal design and construction of the course Web site, accessible to all and making group progress visible to all	
Problems in motivating and structuring collaboration	Choice of a task (communal development of a common product) and instructional strategy (the Jigsaw Methodology, whereby each group member has clear and separate contributions to the group)	Choice of the strategy of combining local and distant members in each group (in this case, Dutch and Finnish students)
Problems in motivating and structuring communication	Strategy of having a rapporteur in each group responsible for responding to the instructor during weekly "on-line activities" structured by the instructor	Use of the mail archive integrated in the course Web site
Problems in maintaining a "group memory"		Use of the shared workspace functionality integrated in the course Web site
Problems in organizing and executing self- and intergroup evaluation	Use of a Web site as the product as well as the process environment allowed convenient opportunities for inter-group comparisons and discussions as well as evaluation based on overall collaborative results	Use of the evaluation tool integrated in the Web site allowed efficient registration of student feedback, immediate comparison among subgroups of students, and immediate visualization to the student of the similarities or differences in his impressions of the course and those of his local and distant coursemates

Table 1: Improving the Efficiency and Effectiveness of Group-Based Project Work: Contributions from the Two Courses

Further Directions

The combination of strengths suggested in Table 1 is being put into practice in our two courses in their 1996 versions. The *On-Line Learning* course, 1996 variant, made extensive use of the BSCW shared workspace and an integrated WWW environment, no longer using FirstClass and a separate mailing system. The 1996 course can be seen at:

http://www.to.utwente.nl/ism/online96/campus.htm

In the 1996 *AIT* course, the communal design of the Web site will be stimulated by applying Jigsaw methodology, also used in the *On-Line Learning* course. Furthermore, communication will be structured by the use of protocols.

We are also actively involved in research issues relating to the support of project work through the combination of appropriate pedagogy and technology. Areas of particular interest in our research include tailoring "group memory" and optimizing communication in terms of both quality and quantity.

References

[Bos, Morgan & Kikstra 1996] Bos, E., Morgan, C., & Kikstra, A. (1996). *Multiple levels of the use of the Web as a learning tool.* Paper presented at ED-MEDIA/ED-TELECOM '96, Boston, 16-21 June.

[Collis 1996] Collis, B. (1996). *Pedagogical re-engineering: Design issues and implementation experiences with the WWW as a learning environment.* Invited Speaker. Paper presented at ED-MEDIA/ED-TELECOM '96, Boston, 16-21 June.

[Collis & Smith 1996] Collis, B., & Smith, C. (1996). *Desktop multimedia conferencing environments to support collaborative distance learning.* CTIT Technical Report Series 96-04, University of Twente, The Netherlands.

[van Diepen & Pouw 1995] Diepen, N. M. van, & Pouw, C. L. M. (1995). The project based course "Applications of Information Technology": Acquiring knowledge and skills using Internet. In W. van Woerden and C. Terlouw (Eds.), *Proceedings, Active and productive learning in higher engineering education* (pp. 87-94). Enschede: University of Twente.

[McManus & Aiken 1995] McManus, M., & Aiken, R. (1995). Using an intelligent tutor to facilitate collaborative learning. In B. Collis & G. Davies (Eds.), *Innovating adult learning through innovative technologies* (pp. 49-64). Amsterdam: North Holland Elsevier.

[Pouw, Terlouw, Joosten & van Diepen 1995] Pouw, C. L. M., Terlouw, C., Joosten, S. M.M., & Diepen, N. M. van (1995). *Tele-project groups in education.* Paper presented at ICTE Orlando '95, Orlando FL.

Explorations: the Cornell Theory Center's Online Science Book

Margaret Corbit
Cornell Theory Center
Cornell University
Ithaca, NY USA
corbitm@tc.cornell.edu

Michael Herzog
Cornell Theory Center
Cornell University
Ithaca, NY USA
herzog@tc.cornell.edu

Abstract: In the summer of 1995, the Cornell Theory Center (CTC) decided to move publication of its printed, periodical, glossy science book to the World Wide Web. The result is the interactive, online publication--Explorations (http://www.tc.cornell.edu/er96/science96/Explorations). In technical design, the book anticipated the availability of new Web browser capabilities and the expansion of graphics capabilities, such as VRML viewers. The development team tried to balance these interactivity-enhancing new technologies with the overriding concern of making the book as appealing and accessible to the audience as possible. Explorations went online in April 1996. We present here a review of the conceptual and technical development of Explorations, along with ideas for extending the capabilities of online publications.

1.0 Introduction: Audience and Scope of Project

As part of its mission, the Cornell Theory Center (CTC), one of four National Science Foundation-supported supercomputing centers in the US, communicates information about research conducted using its resources (hardware, software, and staff expertise) through a variety of media at several levels of technical detail (http://www.tc.cornell.edu). This work includes the production of a periodic book featuring scientific and engineering applications of interest to a lay audience.

CTC's 1996 online science book, Explorations, is intended for a mixed audience that includes researchers, college students, the media, K-12 students, legislators, and the general public. Our goal was to retain the graphic appeal of hardcopy publications and to enhance the publication through the incorporation of hypermedia technologies. We wanted to involve the viewers at the same time that we entertained them.

Target audiences access the World Wide Web via all possible computing platforms, from high-end workstations with accelerated graphics processors to outdated personal computers donated to schools and libraries. In general, CTC tries to accommodate low-end users when providing online information. Since a goal of this project was to experiment with leading-edge technologies for the communication of science, development of Explorations was not constrained in this way. Instead, we attempted to select new technologies and the hardware they would be ported to early on and aimed the technical development at the midlevel machines.

Emerging technologies featured in Explorations include inline animation using a server push, the frames function that was new with Netscape 2.0*, and viewing of 3D files in Virtual Reality Modeling Language (VRML). By designing with frames, we were able to enhance the visual appeal and the graphical navigation of the book. Frames were suppoorted on all platforms at the time this paper was prepared. In contrast, porting of the technology for viewing VRML files has not been consistent among platforms. Acceptable viewers were not available for Macintosh* or UNIX* platforms at this writing. These factors influenced the development process.

Because we wanted to avoid presenting the audience with unnecessary surprises or frustrations, we incorporated a CGI script to detect the reader's browser and its version when he/she first enters the site. This information is presented on a gateway page along with links to sites for downloading the necessary browser and other technologies before entering the publication (these resources are also accessible from within the publication). For VRML, we link to the VRML Repository (http://www.sdsc.edu/vrml), which maintains current technical information as well as access to downloadable software. In addition, we provide a navigation tips screen that graphically and succinctly explains the features of the document and buttonbar functions, a text-based table of contents, and an index-based search capability.

2.0 Technical Development of Explorations

2.1 Implementation of Frames

Explorations is comprised of a series of feature stories describing research in computational science. Through the use of hypertext, the stories are supported with background information, examples from practical applications related to the research, and links to related sites. We incorporated scientific visualizations wherever possible to illustrate the concepts being presented. The structure of the book emerged through a team effort as we explored the possibilities of the frames function.

As one of our first steps after committing to publishing a multimedia science publication on the Web, we created a changeable storyboard using 4x5 file cards, which provided space for a small representation of each screen type and explanatory notes and comments. These aided us in determining the interrelationships of the publication screens within the whole and within a story. Shifting them around and changing their order facilitated planning by trial and error. Once the flow concept was set, the plan was documented and communicated to the team members via a generalized storyboard diagram created in Adobe PageMaker*. The frame structure emerged from this planning process.

Netscape likens frames to window panes in the browser screen, each with a view of a different, but related, part of the site. By clicking in a frame, you can move deeper, either as that particular frame refreshes itself with new files, or as it calls up a file into another frame on the page. In the second model, the original frame remains the same and serves as a reference point. We used both models in designing Explorations, but focused our efforts on exploring the power of the second.

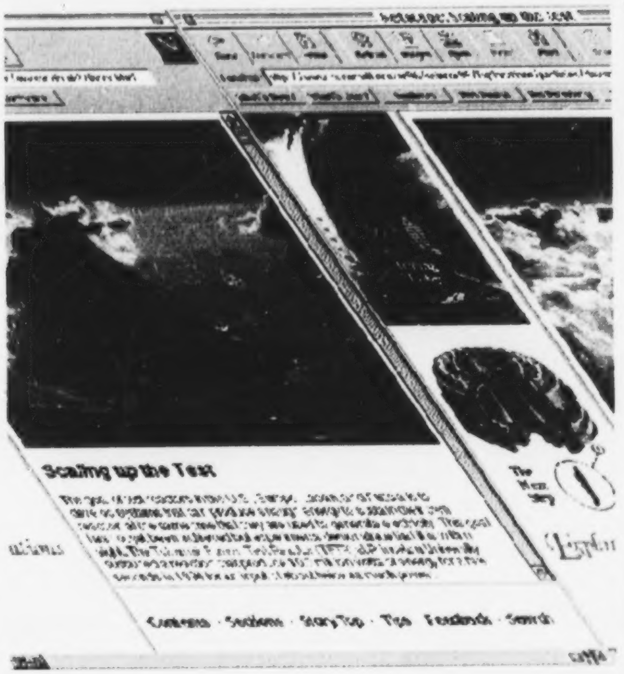


Figure 1. The frames function was used in Explorations as a graphical navigation tool. Here an image map in the left frame holds links to two deeper layers of the story presented in the large frame. The small frame (buttonbar, lower right) gives the viewer access to the whole site as well as help, feedback, and search functions.

At the beginning of the project, we rejected the use of controlled formats such as Adobe Acrobat* or PostScript*. Although they would have provided more design flexibility and artistic control, we saw this as adopting hardcopy technology for online publication, rather than employing new technologies to develop the publication completely in a hypertext environment. We began production of Explorations with the first beta version of Netscape 2.0 (Netscape was estimated to be preferred by approximately three quarters of the browsing audience when we began this project). Cornell University student interns Daniel Cane and Timothy Chi developed the file structure and coding methods used for the Explorations frames.

The use of frames offered several advantages in design and layout of the publication, establishing a "place" for each major component of the pages (graphic image maps, text, illustrations) and, therefore, a consistency of format. Frames also enabled us to maintain a logical and pleasing relationship among the scrolling story, the illustrations, and utility items such as the buttonbar.

All of the top level pages are divided into three frames. On the left, a vertical frame presents a graphic image that is usually a composite of illustrations from the pages at the level below. These images act as graphic sidebars to the pages, enhancing the visual interest at the same time that they present an intuitive graphical guide to the next level in the text. They are sized to fit the initial template and mapped with links to the pages in the hypertext structure to which they refer. Brief text overlays give additional indication of links. The new version of the server software NCSA HTTPd (HTTPd NCSA/1.5.0a), which was installed on our system soon after we started, greatly simplified setting up the image map links.

The major frame for each page is the focal point for that level in the site. Each contains the body of the text, the full illustrations, and often text links to other levels. This is the one frame that we expect to generate scrollbars--since size and font are set by the viewer, we did not attempt to prevent scrollbars here. However, we limited the text length for any segment to no more than three screens in a relatively large font (14 point Times on a Macintosh, "huge" on an SGI). A horizontal frame at the bottom of the

BEST COPY AVAILABLE

page allows the viewer to navigate at a larger scale within the publication. Buttons access the top level of the current story, the entry page for the entire publication, a text-based table of contents, and the navigation tips page.

Any frame on the page will automatically generate scrollbars within its allotted space if the files that are called into it are wider or longer than the space provided. This was inconsistent on different viewing platforms and it was necessary to compromise, determining the final image template sizes by trial and error. (There is an HTML tag that will override the scrollbars, but the results are extremely problematic.) We attempted to minimize the presence of scrollbars on the pages by offering a screen sizing page that functions also as a cover page at the beginning of the site. Viewers are asked to size their browser window to fit the image on this cover page. Because the images in the graphic vertical frames are standardized, scrollbars become necessary only to view oversized illustrations or to scroll text that is larger than the area of the major frame.

2.2 Additional Features

2.2.1 Feedback

We elicit feedback from viewers wherever they may be in Explorations so that we can make continual improvements in our online science publications. A feedback form is readily accessible via the buttonbar, and the messages are sent to the project coordinator as electronic mail.

2.2.2 Search

Explorations is not only an information book, but also serves as an educational and reference tool. We incorporated an ALIWEB* CGI search mechanism, available via the buttonbar, to improve access for students and researchers. This search engine will permit major Web indexed sites to point to the articles in Explorations for the entire World Wide Web audience. Searches are based on the IAFA (Internet Anonymous FTP Archive) template and look for titles, keywords, and descriptions that are provided in the source code for individual pages. This system requires an investment of time by the developers to index the pages manually, but ensures the quality of their accessibility to broad searches.

3.0 Telling the Story with Words: Writing for Hypertext

CTC participated in the development of the National Science Foundation MetaCenter Computational Science Highlights project (http://www.tc.cornell.edu/Research/MetaScience/) during 1994-95, generating new science stories specifically for that site. This experience drove home the importance of breaking apart the text of a traditional, linear feature article into "chunks" of information that are linked and yet stand alone, analogous to, but more complex than, the use of sidebars in conventional print media.

In the hypermedia environment, conventional wisdom suggests that short pages are better. We began with linear features on the research and then took out sections that could become related chunks. Once the story was dissected, it was recast with references to the chunks. Top level pages averaged more than 500 words in length.

Chunks were rewritten to add necessary, though sometimes redundant, information. At this point, chunks often evolved and deepened, especially where we presented practical applications of the research. Chunk length varied from less than 100 to more than 350 words, and this was related to their evolution. The hypertext structure enhances this style of development. By the end of the project, we were writing in hypertext chunks.

While we used images to attract viewers and as lures to deeper levels of information, we also wanted to ensure that people read the text. We therefore kept text and images close together on all the pages, for example interpreting illustrations on the same page as the image. We wrote captions after we incorporated animations and images into the structure of the story to ensure that the context of the illustrations was clear.

4.0 Telling the Story with Pictures: Incorporating Scientific Visualizations and Illustrations

CTC visualization specialists and individual researchers collaborate to present research results graphically, often in three or more dimensions. In some cases, these visualizations represent subjects, such as the human heart, that are easily recognized by a lay audience; in others, the images may be extremely abstract, for example, a solution surface for a complex series of equations. Inevitably, there are aspects of the visualizations that require careful explanation and interpretation. The choice of format is thus important and the complexity of the information included in the images must be considered in terms of clarity as well as file size.

We were able to incorporate still images, animations, and 3D files based on researchers' results to illustrate the stories, although in many instances, these files required reprocessing. In particular, there was a great deal of staff time invested in translating 3D files to the VRML format while controlling for file size and image quality. In addition, most of the animations required extensive editing and reformatting before being included in the site.

4.1 Still Images

We sought to place the research in context by providing examples of related applications. Illustrations for these examples were critically important to the publication, both because the structure required them and because they balanced scientific visualizations which were often exotic in appearance and difficult to interpret. These additional still images were chosen for their ability to add interest and to clarify the stories and were gathered from a variety of sources, both hardcopy and electronic. We found these supporting images almost exclusively via the Web. For example, the image of the Matterhorn was taken from a photograph shot in the summer of 1995 by a mountain climber. We found him by searching the Web using the mountain's name as the keyword and were delighted when he agreed to digitize his best image and then provided the file to us free of charge.

Whereas we enhanced the visual interest of the book by creating the compound images for the vertical frame, we did not modify the actual images to which these were linked so that we ensured accuracy of representation of the researchers' work and fair representation of borrowed images. However, many scientific visualizations were reformatted, resized, or cropped; several were extracted from animations. All were reviewed and approved by the researchers.

4.2 Movies

The animations presented in the book were made available to us in a number of formats ranging from video tape to MPEG movies. Almost all were too lengthy (input video files generated movies of several hundred megabytes), and often too slow-paced to be translated directly into an online animation. We compromised by identifying clips from the tapes that would illustrate concepts presented in the text. We did not attempt to edit any existing soundtracks to fit the clips.

The animations that appear automatically on the pages that present the four sections of the book (for example, Down to Earth) are server pushes. They are generated, image by image, by the server and sent sequentially to the browser. The viewer does not have to wait for the entire file to be downloaded (as much as 25 megabytes) or to store such large files on their machines. When the viewer has a reasonable connection to the network that is not overloaded at the time of viewing, this is a very satisfying option for presenting animations.

With the exception of the four server push files, virtually all the animations were edited in Adobe Premiere* and output as flattened QuickTime* movies (the format required by our UNIX server). In many cases, we adjusted the quality (increasing the contrast during input, for example, to improve the translation from video tape) and compromised on file size through a variety of approaches in order to keep the files as small as possible. Even so, many still fall outside our desired 5 megabyte upper limit for file size.

We chose the QuickTime format for several reasons. First, viewers were available for all platforms. (We make the general UNIX viewer, xanim, available for downloading from our server.) In addition, because there was no simple editing approach for the MPEG format, saving MPEG files as QuickTime movies was the only way to access them without going back to the original digital files (often no longer available). In addition, QuickTime plays more smoothly than MPEG on a Macintosh.

4.3 3D Files

Scientific visualization offers intuitively accessible information about research results and is recognized by many researchers as an important part of the scientific process. In the laboratory--whether it is a physical location with test tubes or an immersive environment in cyberspace--3D visualization of data enables exploratory analyses and aids in processes such as interactive steering through databases. With Web-based technologies, the lay viewer now has the opportunity to fly into and through these files in much the same way as the experienced researcher.

We chose to include a limited number of VRML files in features where exploration of a 3D data set would enhance the information presented and the experience of the viewer. This work depended on conversion of visualization files created for CTC users into VRML format.

The two initial applications approach this goal in different ways. The first, a biomolecular model, affords an experience similar to that of the researcher, presenting a challenge to the viewer in terms of navigating the file. It has an embedded hypertext link that functions as an illustration caption, explaining what the viewer sees. This allows the viewer to explore the environment and learn about the chemistry of the enzyme molecule by clicking on a part of it. We intend to expand the use of this technique in future features.

The second example of 3D illustration, a set of sample files from a geographic information systems (GIS) database of New York State, allows the viewer to fly over selected regions of the state looking at the distribution of vegetation, for example, and is geared more toward entertainment.

To accommodate viewers new to VRML, we grouped the VRML files together with access through an introductory page, Virtual Reality Tours, where introductory information and a link to the VRML repository are provided. The repository maintains up-to-date information on software and formats. Moving from the Virtual Reality Tours page into a model, viewers find a user interface (designed by Cornell University undergraduate Daniel Switkin) in the form of a remote control. At this point, they select a specific model, load it, and fly into it. Viewers exit the models through links in the models themselves or through text links on the Virtual Reality Tours page, which remains visible during their VRML session.

5.0 Self-evaluation and Future Plans

Development of Explorations ended on April 1, 1996 when it was released publically, and while the site has been refined in response to feedback, no major changes have been made to the publication since. Our aim in developing Explorations was to apply the latest available technologies to create a highly interactive, exciting, and information-filled multimedia experience through the World Wide Web. This forced us to explore the limits of such new technology and to develop new capabilities for online documents. This is an ongoing process.

5.1 Overall Design

We experienced an unwelcome design constraint working for cross-platform application. Because of the effect that differences in screen resolution have on the size of the images on the page, we enlarged the design as much as possible to make the publication look reasonable on high resolution screens; this necessitated additional vertical scrolling for low-end Macintosh screens (13" RGB monitors). In the future, we hope that new technologies such as Java* will resolve this problem by allowing the images to be resized on the fly.

This was our first use of frames and we hope that the artificial boundaries they enforced can be overcome though design and a more flexible frame structure similar to the grid system used in conventional publications. Depending on the evolution of Web standards, we may look for alternatives to the frames approach to page design. Access to the broadest possible audience will be an important factor in the decision-making process.

5.2 Text

Five hundred words is probably too much text to include in one piece of hypertext. To further reduce the size of the text pages we will have to rethink our approach to telling the stories. We are also considering a discrete method for providing hypertext definitions--one that will not detract from the flow of the stories for those choosing not to access the definitions.

To enhance the illustration captions, we plan to incorporate audio explanations of the animations and still images into the features. In addition, sound files from scientific data will be included whenever possible and we are currently studying sound file formats for these files

5.3 Illustrations

In this publication, we did not pay a lot of attention to image file size reduction. This will be a priority in upcoming publications. To make this easier, we have acquired Equilibrium Technology's De-Babelizer*. We will also reconsider the use of the JPEG format. We had rejected JPEG in favor of the GIF* format, which permits transparent images, interlacing, and viewing as inline or spawned images on any browser.

VRML and animated illustrations in online publications can be very effective. However, they are not easy to incorporate into the page. Several limitations exist; some will be resolved by advances in Web technology. Others involve careful planning of the publication and close communication with the researchers/visualization specialists. For example, we hope that providing links to sources of information about VRML and download sites for browsers will limit the frustration of those without this capability. However, lumping these illustrations together makes it easier for people only interested in VRML to browse the files and leave without experiencing the rest of the site--for better or worse depending on your point of view. We look forward to the time when inline VRML browsers are commonplace.

6.0 Future Plans

We plan to begin development of our next science publication later this year. With this in mind, we are currently tackling issues such as the potential of Java, JavaScript* or other client-side programming methods and the use of multimedia programs such as Macromedia Shockwave*. One goal is to incorporate small applications into the site as online exercises in computational science. These will allow viewers to explore data sets interactively, for example by watching soil temperature change in response to surface wind speed in a simplified meteorological model.

In the next version we will enhance interactivity by richer use of VRML and possibly through interactive software applets. In addition, there is a potential for the development of interrelated themes within a more interactive publication structure. For example, intermediate pages between treatments of particle physics concepts which appear in both astrophysics and nuclear physics features.

By nature, hypermedia publications depend on the availability of images, animations, sound, and 3D files. As mentioned above, 3D files created by research teams in the process of scientific work did not necessarily suit the editorial and production needs of Explorations. Developing effective illustrations for online publication requires early contact and ongoing communication with the researchers as well as independent resources and expertise to translate and edit files for transfer and presentation on the World Wide Web. The importance of this interaction will only increase as science features become richer and more interactive.

Acknowledgments

Special thanks: We would like to acknowledge the contributions of Linda Callahan, CTC director of external relations, and Dan Dwyer, leader of CTC's Online Information Project, and to thank visualization experts Richard Gillilan, Chris Pelkie, Wayne Lytle, and Bruce Land for their help in many different ways during the course of the project.

This work was conducted as part of the production of the CTC's online science book, Explorations. CTC, one of four high performance computing and communications centers supported by the National Science Foundation, operates a 512-processor IBM SP system. Activities of the Center are also funded by New York State, the Advanced Research Projects Agency, the National Center for Research Resources at the National Institutes of Health, IBM, and other members of CTC's Corporate Partnership Program.

* Trademark or Registered trademark

Designing a Course on the WEB : the Point of View of a Training Institute

Chantal D'Halluin, Stéphane Réthoré, Bruno Vanhille, Claude Viéville
CUEEP - Universite des Sciences et Technologies de Lille
Bat. B6 - Cité Scientifique
59655 Villeneuve d'Ascq Cedex - France
Email. : Chantal.Dhalluin@univ-lille1.fr

Abstract: As an institute of open and distance training, our challenge is to offer to our audience an integrated service (course material, training paths, activities, tuition) then to change information exchanges into learning activities. This paper presents a methodology of courseware development which aims to build a global system of training. To produce the design, we distinguish three spaces : information space, action space and communication space. The shape of the communication space must take into account the available facilities to allow the adaptation of the courseware to the pedagogical practice of each training institute. This is a determining factor in the transferability and usability of a courseware.

Introduction

This paper presents a methodology of courseware development which aims to build a global system of training on the Internet. What attracts the Internet to an educational institute is a large communication network to exchange information in two ways, the on-line browser and the courseware package distribution. So the challenge we have to face is to change information exchanges into learning activities. For this reason, we are interested in second generation servers [Andrews et al. 1995] which respond better to educational needs : better interactivity between video-clip, text, images, and so on ; enabling re-use of all the supports we have developed in a fully integrated manner ; inclusion of graphics and formulae is compulsory for a lot of curricula ; embedded courseware corresponds with the multiplicity of training pathways for individualised training and the ease of navigation required. As a minimum requirement, the system needs communication facilities to enhance real collaboration between users and tutors. In the EONT project, in which we are participating, we are verifying these hypotheses. We have to offer, free of charge, on-line courseware which could be re-used by other institutes, but we must also offer integrated services to our own audience. Moreover, our practice is based on a mix of constructivist and cooperative theory of learning, but the courseware have to support other pedagogical theories and practices in order to be transferable. Therefore to develop our system we distinguish three spaces in which the activities of learners take place : information space, action space and communication space. The communication space depends on the institute, and organises the interactivity between the different spaces to correspond to a pedagogical practice. After a short introduction of educational context of our institute and the content of our project, we characterise the information, action and communication spaces according both technical and pedagogical points of view.

The Educational Context

CUEEP (Centre Universite-Economie d'Education Permanente) is an institute of the University of Sciences and Technologies of Lille in northern France which is concerned with several activities: further education for adults, research into educational engineering (open learning and new communication technologies), transfer within the context of new technologies in education.

The audience of the institute ranges from adults who are unable to count to those who want to get higher education. About 20,000 people attend courses every year, either as a means of improving their general qualifications or to benefit from professional training.

Since the late 70's, we have introduced new learning technologies and important work has been done in designing and experimenting courseware in traditional situations or in distance education and also in research and development of new tools with the help of European programs like DELTA where a system of cooperative learning (the Co-Learn project) [Kaye 1992] has been designed and experimented.

Since the late 80's, we have set up an open and distance system mixing several modes of training (group, individualised, distance, self-training in resource centre [D'Halluin and Vanneste 1995a]). At the moment, people who are registered in distance education are principally those who take a specific examination for adults equivalent to the baccalaureate. They learn from multimedia course material (written paper, audio-tape, video-tape, courseware) and they are in contact with a tutor by phone, fax and Minitel.

Some experiments of the cooperative system Co-learn had been set up during these two last years [Derycke and D'Halluin 1995]. Now we search to integrate this communication system into our distance education organisation.

To continue our work of research into the use of tools of communication in distance education we are setting up a project to design a course on the Web. We have chosen to put on-line a mathematics course whose topic is elementary differential and integral calculus. We have made this choice for several reasons :

This is a compulsory part of the path to scientific higher education. Almost all adults who want access must pass this course - therefore we think that it is important to make it available.

A multimedia package has already been developed (written course, video and software). We plan to use these different support materials within the new course.

The kinds of knowledge are representative of those which the audience of our institute have to acquire when they take a course in our institute. That is to say we find three kinds of knowledge : conceptual and factual knowledge (e.g. the concept of derivation-integration, limits, etc.), procedural knowledge (how to draw the chart of variations in a function ?), strategic knowledge (how to choose the procedure to solve a problem) [Paquette 1994].

Moreover, for about twenty years we have set up an innovative method of teaching/learning mathematics : the "mathematization" of problem-situations [D'Halluin and Poisson 1988] - more a matter of doing mathematics rather than learning mathematics. The learner appropriates him/herself to the content (conceptual or factual knowledge) through mathematical activities (procedural and strategic knowledge).

Six years ago, we faced the following challenge : is it possible in distance learning to develop interactive multimedia course material which : allows the learner to make his/her own way ; give them the initiative to build their own knowledge ; provides the necessary help ; brings synthesis, formalism and the theory required to elaborate academic knowledge [Vanhille 1995] ?

Now in this project we are taking into account the particularities of the new tool (the Web) and we hope that we will be able to meet the following challenge : is it really possible to teach/learn mathematics by doing mathematics using the Web ?

The Methodology for Designing a Course

As the Web is essentially a tool to retrieve information and the interactivity between user and machine is not very easy to create and because we attach much importance to activities of learners in our roles as teachers in charge of a training institute, we have been led to distinguish three spaces to design a course : the information space, the action space, the communication space.

In the next section we describe the features of each space, then we study how each space has to match with the other two to make a coherent global system. We examine these parts in reference to a model of educational situations and learning activities [Derycke and Kaye 1993]. According this model, a situation is seen from different viewpoints (cognitive, organisational, communicational, technical). In particular, the organisational point of view identifies the agents, the roles, the relationships and the resources.

The Information Space

According to the cognitive point of view, the factual and conceptual knowledge takes place in the information space. In our course, they are introduced from a pseudo-real situation (to introduce the derivation-integration concept, the situation is represented by the measures of distance and speed of a car depending on time).

Information about a concept does not progress linearly and is not static : therefore information is given through on-line browsers in a hypermedia mode which may be reached by using languages such as Java or HM-Card [Andrews et al. 1995a] to include a dynamic aspect for animation.

Interactions are forecasted to involve the user. The aim of these interactions is to maintain the attention of the user to help them understand by asking for some details or verification. The user has to remain active. The actions of the user can be characterised by the verbs : to read, to listen to, to look at, to navigate, to answer, to re-do.

By navigating from anchor to anchor, a learner may loose him/herself in the information space. Like any learning activity, the information space should be strongly structured and a learner must know at all times where they are in this global structure. This is possible by using a server like Hyper-G [Andrews et al. 1995]. We need to provide the learner with an automatically updated map to show the path along which they have travelled. With this higher interactivity level, will be the possibility for each leaner to influence the information by annotating or creating new links to personalise their space. To achieve his/her learning objectives, a self-studying learner has to be able, from the information and his interactions, to create his/her own learning activities.

The Action Space

According to the cognitive point of view, the procedural and strategic knowledge takes place in the action space. It is also the space of some meta-knowledge, which is closely linked to the self-assessment.

We can distinguish two kinds of activities proposed in this space : those linked to the content and those about self organisation of learning. They need different tools to be achieved. The activities linked to content are : solving problems (e.g. an optimisation problem) ; training exercises (e.g. drawing a variation table from a graph) ; specifics activities on didactic software (e.g. finding symmetries, translation, etc. which transform a curve into another one) ; and free activities which the learner should initialise by themselves.

They need to use standard office automation software available on the local station , specific files for standard software sent by the server or selected from the information space, transferred packages (an environment composed of software and files which provide the learner with tools, help and results which are useful in resolving a problem and controlling their work) and transferred didactic software linked to a specific objective of learning.

To organise their own training, a learner has to capitalise their work and knowledge by obtaining the results of their activities in the action space. Then they can test their skills and readjust their learning path. Therefore the system has to provide the learner with guidance tools to allow self-assessment and self-control of their course. These will include a hypermedia library which will be filled up by the user, tables of objectives, maps of typical prepared learning paths and an empty map to describe their own route.

After a first analysis, taking into account our academic training context, our opinion is mixed. One of our concerns is that there may be confusion between the information provided by the server and transformations of this primary structure which take place when the user performs actions. Such possibilities could be very useful for the learner to make their own libraries from the information without modifying the server. Another concern is that the two-way transmission between learner and server in the communication space should be as efficient as possible to enhance collaborative learning.

The Communication Space

In order to support an educational process with remote learners, we have to provide communication facilities.

According to the cognitive point of view, strategic knowledge and meta-knowledge take place in the communication space. The actions of a learner can be characterised with three types. The first type is relative to the socialisation : to set up a learning group in the perspective of collaborative work ; the second one is relative to facilitate the pedagogy : to regulate the training, to organise meetings, to exchange, to comment ; the third one is relative to the learning process : to use expertise from tutors or other learners, to debate, to ratify hypotheses, to make decisions, to validate the training.

This communication connects two people (tutor/learner or learner/learner) or a group with or without tutor. Communication can be synchronous or asynchronous.

In our educational practice, we have tested several tools of communication and their uses. We have concluded that the use of several tools is needed to make distance-training successful [D'Halluin 1995b]. According to the use of bi or multi-points, and synchronous or asynchronous communication, their functions are not the same. For example, it is difficult to conclude a scientific debate between learners being in real-time conference, while the report would be produced with an asynchronous conference. The tutor can bring personalised tuition in both modes of communication. Real-time communication facilities like videoconferencing systems have been experimented with during the two last years. We concluded they form no more than 5% of the total communication facilities usage time. The three major reasons are :

- real-time communication puts time constraint on the users;
- real time communication is expensive;
- the communication is rich but not persistent.

Due to these reasons, the current experimentation will only use asynchronous communication in a point to point or group mode.

Point to Point Asynchronous Communication

Electronic mail, by providing an excellent "round trip" time between the tutor and learner, allows more interaction than classical mail ("snail mail") which does not favour a clear separation between the information space and the activity space. In an email system as the messages are stored, they can be referenced and "discussion" can be supported. A negotiation of the condition in which the activities can be done is possible.

Asynchronous communications are cheap and technology is now mature. The institution needs a computer with a mailing system often included in the operating system(e.g. Unix or Windows NT). All that needs to be set up is a group of analogue telephone lines to allow connections from remote learners who only need to add a modem to their personal computers.

Another way of supporting "point to point" communication is to offer annotation on the documents which are on-line. Public or group annotations which can be put on documents from a Web server is an efficient learning tool. Both tutors and learners can add comments to on-line documents.

Efficient systems also have to support "guided tours" of both the information and action spaces. So the tutor can prepare a sequence of work for a specific learner and both tutor and learner can see all the documents which have been "loaded" and read. With this method the learner can discover the current state of a task they have to perform, and the tutor is able to obtain a rapid overview of the learner's work.

These facilities are not yet available on traditional Web servers. Training institutions need to set up specific servers to support these functions. The Hyper-G server developed at the university of Graz already offers these facilities.

From our own experience of the asynchronous communication facilities, we have observed that it still remains difficult for users to support several simultaneous exchanges. The subject field of the email is often the only way to discriminate and to construct a thread of the conversation. In the next section we will continue this discussion as these difficulties will dramatically increase with group communication.

Group Asynchronous Communication

If the tutors encourage collaborative work between learners, the traditional mail is not sufficient. One can use email, but even with the use of list servers to allow multicast within a group, the system do not really help the users to maintain "conversation threads". They use a tree-like classification of messages, by making folders and using sort facilities of the mailing system.

Attachment of documents is possible and standards like MIME allow interoperability between the different client applications which manage the mail.

Dedicated group communication systems have been available for more than twenty years. They provide an automatic way of :

- broadcasting messages amongst a group
- classifying messages in a tree-like structure
- managing conversation threads

Nevertheless, most of these products do not provide a way of structuring the conversation [Hiltz, Turroff 1985] and the computer cannot help the users to get overviews of the group discussion.

In the Co-Learn project, we experimented with a group communication system (based on the speech act theory) which can support collaborative activities within a group. The system offers the user several patterns of conversation dedicated to specific goals. For example, if the tutor needs to negotiate an action (in the action space) with a learner, he selects the appropriate action pattern and he fills in a form which contains the fields dedicated to this negotiation : aim, deadline, etc..

The reader who is interested in obtaining details of experiments into such systems in education should read [Viéville 1995].

Interaction Between the Three Spaces

To distinguish the three spaces we introduced before, is a strategy of courseware development based on transferability. This takes place into the continuum of new learning modes. The first case is face to face learning, where nothing is transferable so the three spaces are indistinguishable. The next case is individualised learning or second generation distance education, for which information and action spaces need mediated supports (papers, audio and video tapes, software), but they are also indistinguishable. Communication modes and collaborative activities are responsibilities of the tutor. However in this case, we encounter a lot of difficulties when transferring our supports to other institutes, since they are dedicated to a particular model of learning. In our case the model is "learning by doing". The institute which wants to integrate our materials into its practice has to agree with this theoretical model, such an agreement is a condition of the transfer.

By distinguishing communication and then information and action spaces, a multiplicity of approaches to the same content are possible : entering by the information space proposes a transmissive practice ; entering by the action space refers to the constructivist learning model ; entering by the communication space between learners enhances a collaborative and social training. Action and information spaces should be re-used by another institute if services are proposed in the communication space. The two way communication between server and client and the possibility to manage author rights at different levels for each component of the initial proposed structure, make it possible for an institute to transform links and structure courseware to its own practice. That means transforming the interactivity between spaces to propose new routes, to enter another objective table, to modify or to enrich the information space in the aim of increasing coherence between spaces and between courseware and practice. What finally makes consistency of the whole are the services proposed by the institute in the communication space. Those services organise the user's learning activities which necessarily develop themselves through the three spaces.

From the learner's point of view, the user often has to move from one space to another to develop his learning activity, and may need to open two spaces simultaneously. For example : the learner needs this flexibility to pick out files or pages from the information space and to transfer them into his own library in the action space ; to research information to solve a problem ; to transfer personal work to other users from the action space through communication space ; to display their objective table or personal map while navigating in the information space ; to discuss at a distance with the tutor about personal work. So must this distinction between spaces be transparent or visible to the learner ? We chose to make it visible . The roles and activities of the user are different in each space. In the first space we identified, they are a reader who receives information, annotates, classifies information, etc.. In the action space, they are an individual learner who solves problems, trains themselves, tests themselves, directs themselves, etc.. In the communication space, they are a social learner who learns academic knowledge with and from others (students and tutors) in a socially organised structure, by debating, by submitting work for the analysis of the others, by proposing hypotheses and so on. Therefore, making the structure visible may help the user to structure his own learning activity which is so important in all individualised or distance learning. A metaphor still needs to be defined to provide the best possible usability of the global system without a long period of training.

Conclusion

In this paper we have dealt with conditions to develop an open and flexible course. Several points have arisen, like the accessibility of resources to a user - some can be free, others require fees. It would be useful to put a virtual tutor on line to guide each learner.

The implementation is evolving from a dedicated solution (Co-Learn project) to a standard solution using second generation systems like Hyper-G which provide a good interface with existing training institution administrative databases with a SQL gateway. Moreover, we are also working to define the interface between the course delivery platform and the production team which also includes teachers.

What we mean by an open and flexible course is a course based on adaptability of available resources, adaptability of communications and a more adaptable structure.

We want the course to be adaptable to different pedagogical strategies (constructivist, transmissive, collaborative and mixtures of them), to different subjects (kinds of knowledge other than mathematics), to different technologies (particularly interoperability problems), to provisions and delivery systems of different institutes. One should be able to measure the efficiency of system in measuring their degree of adaptability.

References

[Derycke and D'Halluin 1995] Derycke, A., D'Halluin, C.(1995). *Co-operative learning in the distance education of adults: why, how and first results from the CO-Learn project*, in Innovative Adult Learning with new Technologies (A-61) B. Collis and G. Davies Editors, IFIP 1995.

[D'Halluin and Poisson 1988] D'Halluin, C., Poisson, D.; (1988) *Une statégie d'enseignement des mathématiques : la mathématisation de situations*, Thèse de doctorat, Lille,France

[Kaye 1992] Kaye, A.(1992). *Identification of Educational Situations,* in *Co-Learn project D2005, Representation models for collaborative educational situations and collaborative learning activities.*

[Paquette 1994] Paquette, G.(1994), *Modélisation des connaissances à des fins de formation*, LICEF, tele-Universite, Quebec.

[Vanhille 1995] Vanhille, B.(1995). *un apprentissage à distance fondé sur la résolution de situations problèmes Entretiens Internationaux EAD*, Futuroscope Poitiers, France.

[Derycke and Kaye 1993] Derycke, A., Kaye, A. (1993). *Participative modelling and design of collaborative distance learning tools in the CO-LEARN project, Teleteaching* (A-29) G. Davies Editor, Trondheim, Norway.

[Andrews et al. 1995] Andrews, K., Nedoumov, A., Scherbakov, N. (1995). *Embedding courseware into the Internet : problems and solutions, Proceedings of Ed-Media 95*, Graz, Austria, pp 69-74.

[D'Halluin 1995b] D'halluin, C. (1995) *La présence à distance, c'est possible ou d'un bon usage des medias, Entretiens Internationaux EAD*, Futuroscope Poitiers, France.

[D'Halluin and Vanneste 1995a] D'halluin, C., Vanneste, A. (1995). *Benefits of a distance mode within an open and flexible system, 17th ICDE*, Birmingham, Vol 2, pp 404-407.

[Hiltz and Turroff 1985] Hiltz, S., Turroff, M. (1985*). Structuring Computer-Mediated Communication Systems to Avoid Information Overload*, Comm. *ACM*, Vol. 28, N°7, pp 680-689.

[Viéville 1995] Viéville C. (1995). *Structuring conversation in asynchronous communication systems to support collaborative learning at a distance*. Telematics for Education and Training, Held & Kugemann.

Teaching Hypertext and Hypermedia through the Web

Paul M.E. de Bra
Department of Mathematics and Computing Science
Eindhoven University of Technology
PO Box 513, 5600 MB Eindhoven
The Netherlands
debra@win.tue.nl

Abstract: Since early 1994 the introductory course 2L670, "Hypermedia Structures and Systems", has been available on World Wide Web, and is an optional part of the curriculum in computing science at the Eindhoven University of Technology. The course has since been completed by more than 200 students from three different universities, two in the Netherlands and one in Belgium. Since January 1st 1996 the course is also offered at two more Dutch universities and at the Dutch Open University.

In order to participate in this course the student only needs a World Wide Web browser. There is no need for separate e-mail, netnews, bulletin boards or ftp software.

In this paper we present the technical environment for the current (1996) edition of the course, which features automatic evaluation of small assignments, a repository for assignment work, a discussion system, and complete monitoring of each student's progress. We also reflect on the previous architecture, used for the first 200 students, and draw lessons from that experience.

1. Introduction

In the spring of 1993 the Department of Computing Science of the Eindhoven University of Technology agreed to add an optional course on hypertext and hypermedia to the curriculum [De Bra 94]. In order to give the students some hands-on experience with hypertext while studying this course we decided to offer the course text in hypertext form, using World Wide Web technology. We also decided to offer the text as a real hyperdocument, not having a linear or strictly hierarchical structure. The URL (or Web address) for the course is http://www.win.tue.nl/2L670/. The course text is written in HTML-2.0, a standard version of the HyperText Markup Language used on World Wide Web. All current WWW browsers understand this version of HTML.

The course places a focus on principles and practice, not on fancy features and multimedia applications. The aim of the hypermedia course is to teach students how to create hyperdocuments that are easy to use, how to develop hyperdocuments from information items and semantic relationships, and how to build hypermedia systems that offer a rich set of useful navigation aids. World Wide Web is the enabling technology, but not the subject matter of the course. Hypermedia functionality not (yet) offered by the Web plays an equally important role as the features offered by popular browsers and servers.

In this paper we focus on the use of World Wide Web technology to make such a course possible, and on the tools we developed and/or used to help both the teacher and the student throughout the course. These tools include tracking student progress, automatic evaluation of intermediate and final knowledge tests, an upload and grading facility for assignment work, and a Web-based discussion system. For remote students without permanent Internet connection (like the students of the Open University) most of the course text also remains available in a static form. Thus, students can download the text and read it off-line to avoid a high phone bill. (In Europe even local phone calls are charged per time unit.)

2. Initial Experience with On-Line Hypertext Courseware

At first the hypermedia course was just an "ordinary" course, taught in a classroom. For such classes, students typically get a book, a reader, or some other kind of hand-out, and they take notes during the lectures. With the hypermedia course the "book" was only available on-line, and the classrooms were not equipped with computers on which the students could make annotations to the pages that were being discussed. In order to give students the opportunity to study the course with a computer, the actual teaching in class was cancelled after four semesters (two in Eindhoven, The Netherlands, and two in Antwerp, Belgium). This immediately made it a lot easier to offer this course to students from other universities. Since January 1996 the course is available to students from six different universities in two countries.

Like many courses at the Eindhoven University this course consists of two parts: the students have to read a substantial amount of information, and they have to complete an assignment to show that they understand and master the principles taught in this course. This second aspect is of course more important than the first.

The hypermedia course deals with the following subjects (chapters):

- definition of hypertext and hypermedia
- history of hypertext and hypermedia systems
- reference models for hypermedia system architecture
- navigation (or browsing) in hypertext
- information retrieval in hypertext
- authoring and converting text to hypertext
- distribution and concurrency issues
- future evolutions

From this list it is clear that there is quite a bit of overlap between subjects. Concepts are illustrated by pointing to the features of systems in the historic overview. Converting text to hypertext involves navigational aspects, and writing hypertext is often done in a distributed way, by means of collaborating authors working concurrently. Because of this (and a lot more) overlap there are many cross-references between subjects, making the course text a real hyperdocument, which is difficult to reduce to a linear (book) structure.

Students have sometimes complained about the lack of a printed version of the course text. For other courses it would probably be better to offer such version, but for the hypermedia course we considered it more important to give the students some hands-on experience with a hyperdocument.

In order to determine whether students have read most or all of the course text, a multiple-choice test is included. The students have to answer 20 questions, by selecting *one* out of four or five possible answers, or for some questions *one or more* out of four or five possible answers. To make it more difficult to solve this quiz by guessing, the student has to answer all 20 questions before getting a score. The only information given to students is the number of correct answers. Every time the student completes all questions, and submits them for automatic evaluation, the score is logged. These logfiles can be used to verify whether a student makes a very large number of attempts to guess the right answers, and also to check whether a student doesn't simply copy the right answers from a friend, and answers all questions correctly on the first or second attempt. The questions are sufficiently difficult and tricky to make it very unlikely that anyone will answer all of them correctly the first time around. In Section 4 [Using Multiple-Choice Tests] we give some suggestions on how to make better use of multiple-choice tests.

The students' grades depend only on the assignment work. Answering all questions of the multiple choice test correctly is just a prerequisite for being allowed to turn in the assignment. This, and the fact that the test is the same for all students, does not encourage them to put a lot of effort in trying to find the right answers in the course text. About 40% of the students copy the answers from a friend (and answer all questions correctly on the first or second attempt). Another 20% is very sloppy and require 50 attempts or more. It appears that these students start from the quiz, and search through the course text for information to get more and more correct answers, instead of reading the complete text before making the first attempt at the multiple-choice test. So far no disciplinary actions have been taken against these students, but the evaluation quiz will be changed as described in Section 4 [Using Multiple-Choice

Tests] to make such abuse less likely.

The assignment for this course is the conversion of a linear document, the *State of the Art Review on Hypermedia Issues And Applications* [Balasubramanian 93] to hypertext. This means dividing the document into fragments of a reasonable size and creating a hierarchical link structure as well as a reasonable number of cross-reference links. (Advice on how to do this can be found throughout the course text.) In order to perform well on this assignment the students need to understand hypertext concepts well, as described in the course text, and they also need to study the document they have to convert to hypertext. (For the assignment a hypertext review is more useful than a document on any random topic.) The students have to encode the hypertext structure by means of HTML, so their hyperdocument can be read using a WWW browser.

In order to grade the assignment, a locally developed tool *www2dot* is used. This tool extracts the link structure from set of HTML documents, and produces a graph, which is visualized by means of the *dot* program from Bell Labs, which is freely available to universities (after signing a license agreement with AT&T) [Koutsofios & North 91] [North & Koutsofios 94]. This graph structure is used as the main basis for the evaluation. It clearly shows the hierarchical structure of the generated document, and the cross reference links (or the lack thereof).

The possibilities for dividing the source document into pieces, and for marking it up using HTML, are almost unlimited. Some students also add special links to indicate the original linear structure, and to point to section and chapter index pages. Although such additions are helpful for the reader of their document, they make it more difficult for the teacher to grasp the real link structure of the hyperdocument the students have generated.

Only one case of fraud has been discovered so far. Two students did most of the work together, resulting in identical structure and HTML markup. They later added some differences, mainly in these extra links to index pages, but not enough differences to not get caught. In Section 5 [Automatic Distribution and Evaluation of Assignment Work] we suggest a way to make it easier to detect cases of fraud.

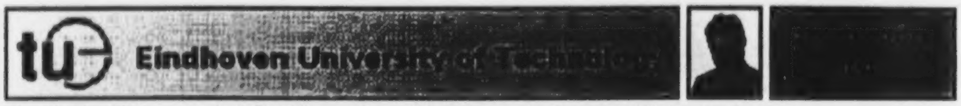
Although this first edition of the course has proven to be usable, and the effort to put the course on line has paid off, mainly because the final grading of a student project can be done in under 15 minutes, a number of problems and suggestions for solutions have emerged over the past two years. In the subsequent sections we discuss these improvements, not only for this course, but for hypertext courseware in general.

3. Tracking Students' Progress

The non-linear, non-hierarchical structure of the course text makes it difficult for students to keep track of their progress, i.e. to decide how many, and which pages they still have to read. The anonymous nature of World Wide Web access implies that this information is not present in the server log files either. The teacher cannot even deduce from the server log how many students are taking the course. Each student completing the assignment is a surprise to the teacher. (Strictly speaking the teacher can already find out about students when they make their first attempt at the multiple-choice test because these attempts are logged separately, with the students' names.)

Another problem students are faced with is distinguishing pages that belong to the course from other documents on World Wide Web. Throughout the course there are links to other relevant information that is not really a *part of* the course. A student has to watch the URL of the document as displayed by the WWW browser to notice leaving the course text.

We have developed a system to track each student's progress. Students now have to register at the beginning of the course. All pages of the course are delivered by a CGI-script, a program that runs on the WWW server, and that keeps track of a student's actions. The student is tracked by supplying pages with the student's identification embedded in the URL (name) of each page. A list of pages that were read, and one of pages still to be read, is part of a standard banner, added to every page of the course. A student can thus easily find out which pages still have to be read, and the generated list contains active links to these pages. Also, the absence of the banner is a visual indication that the student has followed a link to a page outside the course text. We show an example of the standard banner below [Fig. 1].



Paul De Bra read 15/163 nodes. (these read, these still to do.)

Figure 1: Banner of each page of the hypermedia course.

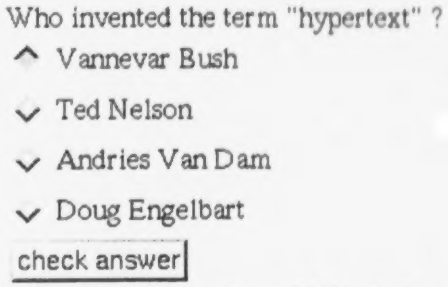
Note that the banner has a link to the home page of the university, the home page of the teacher, which has all address and email information, and to the discussion list, described in Section 6 [A Discussion System on World Wide Web]. The banner indicates how many pages there are in the course, and how many have been read.

Another advantage of keeping track of the student's progress is that it becomes possible to provide context-dependent information. When following a link to a page the student is not yet ready to understand, the system could give advice on which pages to read first, or it could add some extra explanation of terms for which the student has not yet read the definition. Also, superfluous links to pages with definitions which the student has already read can be removed from other pages before presenting them. (The current courseware does not yet take advantage of this possibility.)

4. Using Multiple-Choice Tests

The multiple-choice test in the hypermedia course serves only one purpose: to verify whether a student has read, understood and remembered the course text. The level of understanding that is needed to complete the test successfully may come as a surprise to the student. Therefore, small tests at the end of each chapter are added, to test the student's knowledge throughout the course, and to guide the student to additional information or explanations that may be needed.

To use multiple-choice tests as an aid while studying the course text, it is necessary to offer explanations in case the student gives a wrong answer. (This explanation may contain links to course text pages.) Below is an example of a multiple-choice question [Fig. 2], and a possible reply from the courseware [Fig. 3]. The "final" test with 20 questions, at the end of the course, does not provide such informative feedback.



Who invented the term "hypertext" ?

● Vannevar Bush

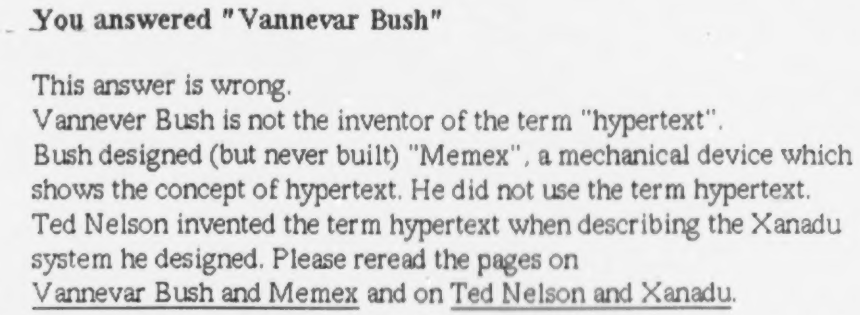
○ Ted Nelson

○ Andries Van Dam

○ Doug Engelbart

Figure 2: Example of a multiple-choice question.

BEST COPY AVAILABLE



You answered "Vannevar Bush"

This answer is wrong.
Vannever Bush is not the inventor of the term "hypertext".
Bush designed (but never built) "Memex", a mechanical device which shows the concept of hypertext. He did not use the term hypertext.
Ted Nelson invented the term hypertext when describing the Xanadu system he designed. Please reread the pages on Vannevar Bush and Memex and on Ted Nelson and Xanadu.

Figure 3: A possible reply to a wrong answer.

For each test the teacher can provide more questions and answers than are presented to the student. The software randomly (based on the student id) picks some questions and answers, and presents them in random order. This makes it harder for students to communicate the correct solutions. In the new version of the courseware the final test with 20 questions is taken from a set of 50 questions.

A final factor in the use of these tests is their influence on the student's grade. Experience with some other courses has shown that students tend to put more effort into tests that count towards their final grade than into tests or assignments that do not count. For the final test the number of attempts and the number of correct answers in each try are counted. For the intermediate tests only the correct answers on the first try are taken into account.

5. Automatic Distribution and Evaluation of Assignment Work

The distribution of the assignment, the actual project work, and turning in the completed assignment all have proved to be problematic:

- The document to be converted to hypertext, and the HTML markup language used for the hyperdocument to be generated offer many possibilities for variation, but do not rule out the possibility of students either producing exactly the same result accidentally, or on purpose by copying. By means of a random generator slight variations can be introduced that are easy to find in the completed assignment. Such variations make it easier to detect fraud (by examining the generated document), without having to keep all projects of all students for many years (for comparison with new project work).
- The project work requires many skills, related to authoring hypertext, such as selecting an appropriate page size, creating an appropriate structure of hierarchical and cross-reference links, adding links to the list of bibliographical references, etc. The most common problem with the assignment is that about 15% of the students "forget" to create cross-reference links, thus generating a purely hierarchical link structure. By means of small assignments, after each chapter, the required skills can be learned more or less separately, before starting the final project. Such small assignments, like generating a graph simulating a hyperdocument's structure, can be easily graded automatically. A small assignment can teach students that for this course a purely hierarchical structure is not acceptable. They can then realize they have to create cross-reference links before completing the final project.
- Handing in the project has been the biggest problem in the past. The students in Eindhoven and Antwerp have accounts on a Unix machine. The assignment work can be easily copied from their home directory. Unfortunately, many students fail to set the file permissions right so the teacher can access their work while their fellow students cannot. (Many students block access to everyone, including the teacher.) For both the smaller assignments and the final project, the DReSS system can now be used [De Bra & Aerts 95]. This "Document Repository Service Station" makes it possible to upload documents or even complete directories to the department's Web server by means of a form that is part of the course text.

DReSS is similar to the well known BSCW system [Bentley et al. 95]. (BSCW stands for "Basic

Support for Collaborative Work".) DReSS is much simpler than BSCW. It only provides upload and modification facilities, with full authorization facilities, but no version control or workspace overviews. Some newer browsers support file upload. DReSS uses a different upload method, with a small helper application that can be downloaded from the course text, and that is very easy to install. With DReSS students can upload a whole directory at once. (With BSCW or the browser-supported file upload this would not be possible.) The advantage of using DReSS instead of direct file access to the students' directories is that, after uploading an assignment, the work can be checked using CGI-scripts, because it becomes part of the Web server's document space.

6. A Discussion System on World Wide Web

Students from different universities almost never meet in person. A discussion platform enables them to exchange ideas and remarks on the course. Instead of using a mailing list, which would require the students to get acquainted with e-mail, and mailing list archives, we have selected the WWW-based Futplex system [Holtman 95], developed at the Eindhoven University. Futplex has been used in several projects and Web sites already, including the Digital City of Amsterdam.

Futplex lets students talk publicly by posting messages and replying to others' messages, and they can create new pages, representing new lines of discussion. Each time a student enters the Futplex system, an index page to all recently added messages is presented. Futplex offers most of the functionality of Internet Netnews, but does so through the Web. It allows special "chat" pages, for discussions of a very temporary nature, and it also allows "shouting" to quickly get in contact with another student.

7. Conclusions, Advice and Future Work

7.1 Conclusions and Advice

World Wide Web offers interesting possibilities for creating on-line courses. Our experience with four semesters of teaching the course 2L670 on "Hypermedia Structures and Systems" for students from three universities, and with adapting the course for students from two other universities and the Open University, leads to the following conclusions and remarks:

- Not having a paper version of the static portion of the course text is a disadvantage. For most courses it should be possible to offer a printout. (For the hypertext course we explicitly wanted a hypertext document.)
- The anonymous nature of World Wide Web makes it difficult (for the teacher and the students) to find out what each student has read. Our new tracking software solves this problem by generating personalized logging information.
- On World Wide Web everything is connected. We have provided a uniform banner to distinguish pages of the course text from other Web documents.
- Students can be helped by intermediate tests and assignments, in addition to a final test and project. Small projects can easily be evaluated automatically.
- Multiple-choice tests should be varied automatically to prevent copying.
- The DReSS system enables students to hand in their assignment work through the WWW browser, and allows verification of the work by means of CGI-scripts on the WWW server.
- The Futplex system lets students from different universities discuss the course with each other, and with the teacher. It only uses the WWW browser and server, so it does not require setting up email software and mailing lists.

7.2 Future Work

In the future we will use DReSS and Futplex to enable students from different universities to not only discuss the course but work on the project together.

We also plan upgrade the course text to HTML-3.2, to make the presentation more attractive by means of tables and text flow around figures. We will move some functionality to the browser by means of client-side imagemaps, to reduce the server load. We also plan to include dynamic illustrations using Java. We wish to keep most of the interaction server-based however, in order to remain able to do logging of each student's progress.

More universities and other institutes for higher education have expressed interest in offering this course to their students. Further automation of the evaluation and grading of assignment work will be needed to cope with the increasing number of students taking the course.

8. References

[Balasubramanian 93] Balasubramanian, V., (1993). State of the Art Review on Hypermedia Issues And Applications. Independent Research Report part of PH.D. Program, Graduate School of Management, Rutgers University, Newark, New Jersey. Converted to HTML by Denys Duchier. (http://www.isg.sfu.ca/~duchier/misc/hypertext_review/).

[Bentley et al. 95] Bentley, R., Horstmann, T., Sikkel, K., Trevor, J. (1995). Supporting collaborative information sharing with the World Wide Web: The BSCW Shared Workspace system. Fourth International WWW Conference, Boston, WWW Journal, 1, 63-74. (http://www.w3.org/pub/WWW/Journal/1/bentley.151/paper/151.html)

[De Bra 94] De Bra, P. (1994) Hypermedia Structures and Systems. Course 2L670, Eindhoven University of Technology. (http://www.win.tue.nl/2L670/)

[De Bra & Aerts 95] De Bra, P., Aerts, A. (1995). Multi-User Publishing in the Web: DReSS, A Document Repository Service Station. NLUUG Autumn Conference, "Publishing on the World Wide Web", Ede, 13-26. (http://www.win.tue.nl/~debra/dress/dress-paper.html)

[Holtman 95] Holtman, K. (1995). The Futplex System. ERCIM Workshop on CSCW and the Web, GMD, Sankt Augustin, 25-30. (http://orgwis.gmd.de/W4G/proceedings/futplex.html)

[Koutsofios & North 91] Koutsofios, E., North, S.C. (1991). Drawing graphs with dot. AT&T Bell Laboratories, Murray Hill, 59113-910904-08TM. (ftp://ftp.research.att.com/dist/drawdag/dotdoc.ps.Z)

[North & Koutsofios 94] North, S.C., Koutsofios, E. (1994). Applications of Graph Visualization. AT&T Bell Labs. (ftp://ftp.research.att.com/dist/drawdag/koutsofios-north-GI94.ps.gz)

MRTspace - Multi-User 3D Environments using VRML

Dieter W. Fellner, Oliver Jucknath
Department of Computer Science, University of Bonn
fellner|jucknath@graphics.cs.uni-bonn.de

Abstract: The paper presents an architecture for multi-user interaction in 3D worlds which is based on existing server technology to handle the communication aspect and on VRML (and its recent extensions)

as the presentation level protocol. To implement *multi-user environments* we propose the use of the HTTP (Hypertext Transfer Protocol) to share the static definition of a common room and the IRC (Internet Relay Chat Protocol) to exchange the dynamic Information between the participants.

Enhanced realism can be achieved by providing physical continuity across WWW links (eg. geometrically seamless passing through doors), collision detection, solid walls, intuitive movements in 3D environments including the climbing of stairs.

Our experiences are based on a prototype implementation called *MRTSpace*, a stand-alone VRML browser which was designed as a platform to implement experimental extensions to the VRML standard.

Keywords: VRML, IRC, multi user environment, MRT

- Introduction
- Improving the 3D perception in VRML worlds
  - Navigating along solid objects
  - Climbing Stairs
  - Geometric and visual `continuity' across Rooms
- Adding Multi User Support
  - A new Node for VRML
  - A new Protocol for IRC
- Conclusion
- References
- Acknowledgment

Introduction

Two key points for bringing VRML closer towards the vision of a cyberspace are a more realistic behavior of the VRML objects together with a multi-user support.

The major problem with most current VRML browsers is the relatively low level of navigation support either requiring a long training period or resulting in users frequently getting lost in their 3D scene. Users very easily loose orientation, either by stepping `into' an object (e.g. diving into the ground or ending up between two sides of a wall) thus making the reset-to-original-viewpoint button the most frequently used one or by the rather sudden change of the environment when following WWW links.

To give a more realistic feeling of the scene MRTSpace has several levels of collision detection which directly affect the path of the camera and thus meet the user's intuitive expectation of solid objects. Controlling the user's camera path by methods of distance control and collision detection further introduces a consistent way to `walk' on the ground (i.e. keeping the distance to the ground at a constant value), including the climbing of stairs of the jumping into ditches.

Furthermore MRTSpace has several heuristics to determine whether a WWW link leads to a continuation of the current scene, e.g. an adjacent room. In this case MRTSpace loads the new scene, merges it with the current scene and allows the user a smooth transition from one room into the next. This way, the user can keep track of his or her position in 3D space, which is one of most important reasons to use a 3D world for navigating the WWW.

Another quite unique feature of MRTSpace is the support of multiple users sharing a single room, with several ways of interaction, thus leading to the original concept of cyberspace as a means of communication. Thus MRTSpace merges two technologies: IRC [OR93] as a medium to communicate and VRML [BPP95a] as a interface to a virtual reality. IRC was originally developed to allow people all over the world to chat with each other. It has been used to develop a network of servers which form the backbone of a very effective and fault tolerant multi user medium. Using the HTTP for sharing the constant data of a room where people can meet and using IRC to exchange information between these users avoids several start up problems that arise when introducing new specialized servers.

MRTSpace is based on *MRT*, a Minimal Rendering Tool [Fel96], developed at the University of Bonn. MRTSpace builds on the modeling functionality and the structure of geometric objects maintained by MRT to represent all of its internal data, handle the interaction between the different actors and do all the visibility computations, including collision detection. This toolkit also resolves some of the major problems of developing a platform independent program: it has a platform independent graphical user interface which uses most of the existing graphics hardware to display 3D objects.

Improving the 3D perception in VRML worlds

Navigating along solid objects

A general problem of 3D viewers is the handling of solid objects (like walls) when the user walks through a scene. The minimum level of user support is the detection of solid obstacles and the prevention of moving the user's camera into 'black holes', i.e. 3D areas surrounded by non-transparent planes thus blocking all light sources and thus resulting in a completely black screen. Some viewers provide a *head-light* feature to circumvent this problem of getting lost in the dark. Nevertheless, light is not the only issue as the user is still confronted with a uniformly colored screen.

A better navigation control can be built on functional elements commonly used for ray-tracing algorithms [Gla89]: most of the time in the ray-tracing process is spent on casting rays onto the scene and determining which is the closest object. The considerable research invested in ray acceleration techniques provides us with a rich set of tools (e.g. octree or regular space subdivision) which can be applied to various aspects of navigation control.

MRTSpace uses the object-oriented data structure of MRT to represent all VRML objects. Consequently, collisions can be detected by using the ray-tracing functionality of MRT. The built-in ray-tracing modules compute the intersection of a given ray with a scene, returning the distance to the closest object hit.

So if the user moves into a given direction MRTSpace casts a ray along the trajectory and checks whether the returned distance is less than a given value. In this case the user was about to run `into' an object and MRTSpace takes the normal vector of this object at the intersection point of the trajectory and the object to determine how to change the user's direction in order to avoid a collision or disappearance in `outer space'. This feature has the advantage that a user who is too close to an obstacle (assuming the starting position of the camera has been chosen inappropriately by the designer of this scene), and therefore sees nothing but a solid color, just needs to go ahead and will automatically be repositioned at a more appropriate point.

Figure 1: Using rays cast ahead of the moving avatar to guide the user

Implementing this feature in MRT can be done with the following code (which is much shorter than its description):

```
// find closest object 'hitObject' intersecting the ray emanating from
// 'eyePoint' in direction 'dir' returning the distance to the intersected
// object in variable 'dist'. In case that no intersection occurs
// if 'hit' equals FALSE all other variables remain unchanged.

hit = allObjs->intersect(eyePoint,dir,NULL,dist,hitObject);
if (hit && dist-HEAD_DISTANCE<stepwidth)
{
   // assign normal vector in intersection point to 'norm'
   t_3DVector norm=hitObject->surfaceNormal(eyePoint+dir*dist);
   // change user's direction accordingly
   if ( dot(dir*norm,up) < 0 )
      TurnRight();
   else
      TurnLeft();
}
```

Climbing Stairs

Another problem frequently encountered with 3D viewers is the inadequate support for navigating over uneven terrain or for climbing stairs. Having taken care of the collision problem moving on level ground can be mastered by most users without significant problems. But navigating on a slope or climbing stairs adds an extra degree of navigational complexity and typically gives the user a hard time.

An easy solution of keeping the user at a constant distance over the ground is implemented in MRTSpace by casting two rays down to the ground: one ray before and one ray after each step. Both intersections return a distance. If the difference of the distances is zero the user is considered to walk on flat ground and no special action is taken. If the difference is within a small range the user is considered to *take a step* and the difference is added to the position of the user. If the difference is negative and within a rather large range the user is considered to *fall* and the position of the user is decremented in several steps to the new level, giving the user the impression of a fall. If the distance of the new position towards the ground is too large, the user is either set to the *fly mode* or just stopped depending on some

BEST COPY AVAILABLE

setup parameter. If the user is in fly mode, the casting of rays towards the ground is used to automatically switch back to the *walk mode* when the user *lands* on the ground.

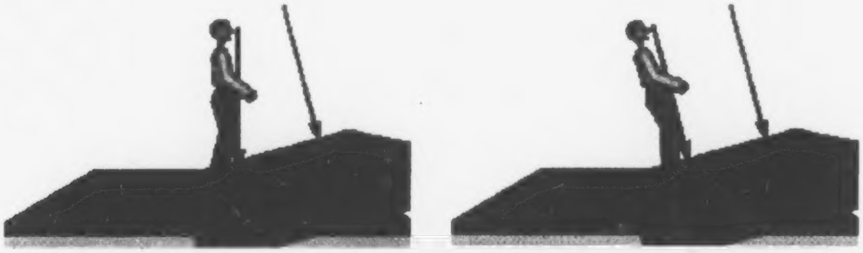


Figure 2: Adjusting the up vector of the avatar to the normal of the walking plane.

Once the scene structure is consistently exploited staying *upright* is not too difficult either. In this mode the viewer casts a ray down towards the object the user is moving on and calculates the angle between the surface and the up vector. If this angle exceeds a given threshold value the up vector is continuously set to the normal vector to the ground position. The most impressive example in terms of user comfort for this feature is a big ball representing a planet where the user has the ability to walk on this ball, the feet always staying on the ground and the gaze vector always parallel to the tangent plane.

Geometric and visual `continuity' across Rooms

A completely new approach is the concept of *visual continuity* across (adjacent) rooms. If the user selects a WWW anchor, MRTSpace starts a child process that retrieves the description of the new location. When the child returns and the WWW link led to a HTML page the viewer starts an external browser. But if the web link led to another VRML scene MRTSpace tries to find a door between the new and the old scene. A *door* is an object that has web anchors leading from one scene to another and vice versa. If a door is found MRTSpace tries to match both scenes by matching the doors of each file. In the trivial case the 3D representations of both doors are identical, so the match can be done in an exact way. If the shapes of the doors differ, the bounding volumes are used instead and some heuristics try to figure out the direction of the doors. In both cases the current and the new scene are merged, the door object becomes penetrable and the user can walk from one scene to the other without any visual interruption. This greatly enhances the perception of the overall VR environment as a place to wander around.

Comfort of navigation can further be improved by two additional features: *load ahead* and *keep rooms*. Load ahead means that MRTSpace tries to load all WWW anchors that lead to a URL matching something like `*.wrl`, searches for doors, and adds these rooms to the current scene whenever possible. So if the user enters a room with several doors, these doors will `open' as soon as the representation of their web anchors have been loaded. The Keep Rooms option controls when to discard the geometric description of rooms. In the simplest case this is done immediately after the user leaves, but if the resources of the machine that runs the browser aren't too limited is very convenient to keep at least the last room visible.

Adding Multi User Support

Recent proposals on behavior all try to bring VRML worlds closer to the notion of cyberspace as a means of communication [Bro95,BE95]. Most proposals, however, follow the route of defining a new server architecture to take care of the exchange and interaction aspects involved in multi-user virtual reality. Others like [MZP*95] are based on the multicast backbone approach MBone [MB94].

In contrast, our approach builds on existing server technologies merging IRC as a medium to communicate and VRML as a interface to a virtual reality. IRC was originally developed to allow people all over the world to chat with each other. It has been used to develop a network of servers which form the backbone of a very effective and fault tolerant multi user medium. Using the HTTP client/server protocol for sharing the static data of a room where people can meet and using the IRC protocol to exchange information between these users avoids several start up problems that arise when introducing new specialized servers.

A new Node for VRML

For our solution we need a simple extension to the VRML standard of the following form:

```
SharedRoom {
   width   2  # SFFloat
   height  2  # SFFloat
   depth   2  # SFFloat
   name    "" # SFString
}
```

The first three fields describe a bounding volume. The VRML browser will check each time the user moves (and on start up), whether the user enters one of these bounding volumes. In this case the VRML browser opens a connection to the next IRC server and uses the IRC command `JOIN` to open a channel with the name given in the fourth field. This name must be unique for this shared room. One obvious way to build such a name is to choose the URL of this VRML script and append some extension to identify `SharedRoom` within the script. The channel with this name is then shared among people all over the world who have loaded this particular script and entered this particular room. If the channel did not already exist the IRC server will create it automatically.

A new Protocol for IRC

The protocol used by the VRML browsers to exchange information needs to be embedded within the standard communication allowed by the IRC protocol. Thus all the people in `SharedRoom` can use the facilities of the IRC: they can talk to all other participants, they can directly talk to each other on private lines and perform personal interaction. Furthermore, there is a special protocol of VRML messages, which contains the icon description and the movements and actions of a user. These VRML messages are sent using a standard IRC `PRIVMSG` command, but they are marked as VRML messages by a starting exclamation mark ('!') followed by a three digit code. (If the user enters a message starting with a '!' it should be sent as "!!", with the second '!' removed by the receiving VRML browser).

To implement our proposal, only the following codes need to be implemented:

```
!000 Sign on message
!001 Icon definition
!002 Position definition
!003 Icon redefinition
!004 Position redefinition (movement)
!005 Log out message
!006 Click on icon
```

All these messages are of the form:

```
!nnn[<VRML script>]
```

For example:

```
"!000Separator {Cube{}}"
"!005"
```

Multiple VRML messages within one IRC message are allowed, each starting with a `!nnn`. If a user enters a channel, the viewer sends a `!000` followed by a VRML Separator containing the description of his icon. As a response to this all other users which are already on this channel send a `!001` followed by a VRML Separator containing the description of their icons and a `!002` followed by a VRML Transform which describes their absolute position and orientation. This way the VRML browser gets all the information it needs to visualize the other participants of this shared room. A moving participant will send a `!004` followed by a VRML Transform, which contains the new position and orientation of the user. The IRC protocol provides the necessary information to assign this movement to the corresponding icon. It is up to the VRML browser to prevent the icon from popping around by smoothly interpolating between successive positions.

Code `!003` is used to redefine the icon of a user. This may be used to express some (minimal) gesture by changing the icon according to the mood of the user. If a user clicks on an icon representing another participant, a `!006` message with no parameters will be sent to him. Finally the VRML browser will send a `!005` if the user leaves the bounding volume of this `SharedRoom`.

There are two special cases for the definition of icons: It is allowed to use the DEF/USE mechanism of VRML to avoid resending the icon of a user, so this would be a legal sequence:

```
"!000uni-bonn.de:DEF std   Separator {Cube  {}}"
"!003uni-bonn.de:DEF smile Separator {Sphere{}}"
"!003uni-bonn.de:USE std"
```

Another case is the usage of `WWWInline`, which should be allowed to send a description which exceeds the maximum line length of 511 for an IRC message.

To summarize, the advantages of choosing IRC are:

1. Efficiency: IRC has an established network of IRC servers, which build a backbone for broadcasting the messages of each IRC client around the world. Combination techniques bundle many small messages into larger packets by which the IRC network avoids sending a message twice over the net. For example imagine two users A1 and A2 in the USA and two other users G1 and G2 in Germany. If A1 sends a message it is sent over the ocean once by the american IRC server to the german server which in turn distributes it to G1 and G2. Theoretically the same effect could be achieved by an IP Multicast message [MB94], but actually IP Multicast messages do not work over long distances, because most routers are configured to filter out multicast messages.

2. Fault tolerance: If one IRC server is down switching to a different one is done automatically. The IRC network itself has several mechanisms to cope with network interrupts, unreachable servers and other real world trouble.

3. Easy startup: No special service providers or new server programs are necessary. Start up can be done by simply providing a new VRML browser and using existing IRC servers. Even if there is a need to create some new servers, providers will be much more willing to install an IRC server, which uses technology that exists for quite a while and is well tested, than a brand new experimental server with unknown safety hazards.

The only drawbacks using the IRC to implement a multi user virtual reality are the limited message length (messages are limited to 511 bytes), the delays introduces by IRC servers with enabled `flood control' which allows only one message per 2 seconds, and the difficulties in implementing semaphores

BEST COPY AVAILABLE

(although it can be done).

Conclusion

We have presented a proposal to introduce multi-user interaction into VRML worlds which is based on existing IRC server technology to handle the communication aspect and on VRML as the presentation level protocol.

The implementation of this approach together with several new features improving the way users can navigate through threedimensional scenes is available through *MRTSpace*, a stand-alone VRML browser built as a platform to implement extensions to the VRML standard and to experiment with new features of user interaction.

References

BE95 Broll W., England D.: Bringing worlds together: Adding multi-user support to VRML. In *VRML'95 - First Annual Symposium on the Virtual Reality Modeling Language* (1995), ACM SIGGRAPH, pp. 87-94.

BPP95a
Bell G., Parisi A., Pesce M.: The virtual reality modeling language, version 1.0 specification. `http://www.sdsc.edu/SDSC/Partners/vrml/Archives/vrml10-3.html`, May 1995.

BPP*95b
Bell G., Parisi A., Pesce M., Mitra, HArdenbergh J.C., Meyer T., Martin B., Carey R., Marbry J., Blau B.: The virtual reality modeling language, version 1.1 draft. `http://vag.vrml.org/vrml-1.1.html`, Dec. 1995.

Bro95
Broll W.: VRML: From the web to interactive multi-user virtual reality. In *Modeling - Virtual Worlds - Distributed Graphics*, Fellner D. W., (Ed.). infix, Bonn, 1995, pp. 191-200.

Fel96 Fellner D. W.: Extensible image synthesis. In *Object-Oriented and Mixed Paradigms*, Wisskirchen P., (Ed.), Focus on Computer Graphics. Springer, Feb. 1996, pp. 1-18.

Gla89
Glassner A.S. (Ed.): *An Introduction to Ray Tracing*. Academic Press, London, 1989.

MB94
Macedonia M. R., Brutzman D. P.: MBone provides audio and video across the internet. *IEEE Computer* (Apr. 1994), 30-36.

MZP*95
Macedonia M.R., Zyda M.J., Pratt D.R., Brutzman D.P., Barham P.T.: Exploiting reality with multicast groups: A network architecture for large-scale virtual environments. *IEEE CG&A 15*, 5 (Sept. 1995), 38-45.

OR93
Oikarinen J., Reed D.: Request for comments: 1459, Internet Relay Chat Protocol (IRC).

BEST COPY AVAILABLE

ftp://ftp.dfn.de/pub/doc/rfc/rfc-1400-1499/rfc1459.Z , May 1993.

Acknowledgment

We are grateful to all our colleagues and students contributing to the development of MRT which is the result of a collaborative effort in the Computer Graphics Lab of the University of Bonn (see http://hyperg.cs.uni-bonn.de/CompGraph). This particular work builds extensively on the contributions by H. Bendels, M. Fischer, A. Hopp, G. Müller, and S. Schäfer.

Delivering *The Daily Us*

Barry Fenn
HyperMedia Unit
The University of Auckland
Aotearoa New Zealand
barry@cs.auckland.ac.nz

Jenny Shearer
Political Science Department
The University of Auckland
Aotearoa New Zealand
jshearer@cs.auckland.ac.nz

Abstract: We have a while longer to wait for the computer-based intelligent agents that can search out, evaluate, and deliver personalised selections of global media resources to us on a regular basis. An opportunity exists right now, however, to implement network environments that distribute and prioritise hypermedia 'suggestions' offered within communities of human associates. Such a ëDaily Us' would incorporate layers of automated selection and prioritisation of references sent and received among individuals and interest groups. This same sort of media interface will then naturally come to assimilate whatever resources are eventually gathered by intelligent agents. The result will be a system serving the personal entertainment and information needs of each user, while also incorporating and enhancing important social integration functions now served by print and broadcast media. Some of the interactive structures needed to effectively coordinate such overlapping patterns of cross-reference may then evolve to support fundamentally new modes of human communication.

Introduction

Most people in ëdeveloped' societies already expect to access a wide variety of media coming from sources distributed around the globe. In fact, most people's lives have become very closely intertwined with existing ëchannels' of access to media. Selection of what is important or worthwhile, however, still usually depends in the first instance upon a relatively few ëgatekeepers' such as editors or professional reviewers. The next major cultural epoch will see us all empowered with the means to powerfully influence *each other's* selection process.

The ideal Web interface will fully acknowledge not just the user, but the user as part of an extended community. Each individual Internet user's community associations are unique and now becoming increasingly independent of geographic proximity. A substantial part of the social conversation that sustains such online communities revolves around shared experience of particular media reports or resources. In turn, whatever specific blend of media a person happens to encounter exerts a strong influence on their whole pattern of individual identity - directly affecting their perceived relationship with the world at large. All of this represents a domain of potential interactive enrichment that needs to be supported *much* more intentionally.

Usage of the Internet also reveals a strong user interest in various aspects of ëcommunity' in its own right, as well as personalisation. The two domains of information are not separate issues; when integrated, they develop a social utility. That is, the individual is assisted in carrying out whatever tasks or recreation he or she has in mind, while also being advised of information they might "need to know" for purposes of the common good - one of the traditional functions of mass media.

ëThe Daily Us'

The move toward completely user-defined media environments is often criticised for neglecting 'serendipity' - the chance to accidentally discover important information - by limiting the user to only what they might expect to encounter. This is an important criticism, but one that overlooks the dominant trend toward more and more widespread interpersonal 'sharing' of media resources that has been steadily accelerating since photocopiers came into common use. A number of electronic utilities serving this sort of function already exist on an ad hoc basis (newsgroups, email nicknames, lists of 'cool sites', and such-like), but we have not been able to locate any documented attempts to consciously engineer this convergence - nor even any clear suggestions about the general way in which such a seemingly obvious use of computer technology might be implemented!

Individuals will soon be able to define the behaviour of their own personal information system, based on input from a wide range of other people and drawn from many sources - a far journey from the days when people opened their daily paper, read the contents, and considered themselves to be well-informed. On the other hand, the general concept of a "newspaper" - with size and location defining the relative importance of content - still provides the best metaphor to use in discussing something that otherwise is essentially brand new.

For more than a decade now we've been awaiting the "Daily Me", a digital newspaper concept originally put forward by Nicholas Negroponte and elaborated by many others. Such an integrated service is envisioned as offering a personalised selection of information that has all been actively located and filtered by intelligent agents, based on criteria set by the user according to their particular interests. Negroponte also makes a tongue-in-cheek reference to the "Daily Us", but just as a rather marginally-functional agent of such personalisation; having a filter than may be adjusted only according to one's mood or time available for mainstream "features" type material. He goes on to describe a more astute "digital sister-in-law"; an interface agent with stored knowledge of both one's expressed media preferences and actual past responses [Negroponte 1995]. All this assumes, of course, that most of the information relevant to a particular person's life can be automatically gathered together from the various roadside attractions to be found scattered along an information superhighway.

We contend that a key factor is generally being undervalued in all of this theorising about the future potentials of computer ëintelligence' to automate online media selection - that is, the existence of a vital community of intelligent humans - and that ëThe Daily Us' is in fact an apt name for a powerful class of application that could be made available here and now. To implement this we need to develop effective support for a continuous feed of relevant references offered by any member of the community - which are received unsolicited by the user - so that these can then be used in the automatic compilation of a unique final presentation.

Even when we eventually have intelligent agents which are able to serve Negroponte's "digital sister-in-law" function, the recommendations of other real people will continue to fulfil a unique function - particularly in terms of exposing us to critical pieces of information from areas we'd never think to intentionally explore otherwise. Most importantly, though, we don't have to wait for more powerful systems in order to share the results of our own innate intelligence much more effectively. Hyperlinked suggestions received from a user's friends, professional colleagues and other circles of mental association can already be very effectively fielded using the capabilities of existing personal computers.

There are a few existing initiatives which can provide some insight into the range of considerations involved in letting other people help structure a user's access to information. The most relevant is probably the NoCeM project, intended to perfect selective controls for access to Usenet files. Though primarily directed at blocking out unwanted postings from view, it can also work the other way around, to only acknowledge postings from specifically authorised persons - or specific postings that those other ëapproved' people have recommended [ëmoose' 1996]. Beyond this, there is a body of work that has been done within the field of Computer Supported Collaborative Work that will provide useful hints toward designing an appropriate user interface.

Process of Interaction

The way in which suggestions are filtered provides a good starting point from which to discuss the proposed interactive infrastructure which could eventually lead to a highly-personalised version of ëThe Daily Us': Anyone could recommend anything to anyone else; without being concerned that this might be taken as an unwanted intrusion. Such recommendations would be automatically filtered and ëgraded' as to their probable importance (i.e., whether or not they appear on the "front page") based on an process which has assigned stature to whatever past suggestions they may have made - all by way of one's own personal assessment of the media itself. In short, whoever proved to be the best source of references would naturally come be promoted to the most effective ëvirtual editor' position.

The likely ërelevance level' of known contributors would initially be assigned by the user, but the stature of each referee would then tend to change adaptively over time. (Clearly such an assessment of others' judgment would usually be treated as highly personal and thus held only in encrypted form until needed.) This process could be made visible to the user, or left to function purely in terms of how appropriate any particular media resource was eventually found to be. Of course, whoever forwards any particular suggestion could also specify their own assessment as to its priority, which would be automatically factored in to the overall process of evaluation.

Thus the recommendations of a good friend, or some reference that was designated by a well-respected professional colleague as being particularly important, might automatically be accorded "headline" status; at least in some relevant section of content. Likewise, if a significant number of the 'lower' level members of one's chosen community recommend a particular piece of information it will be automatically promoted to a relatively more important position. Regular reference contributors might be structured into subject-based interest groups, with all those on a physically localised freenet being one natural example. In fact, it seems likely that whole new patterns of online association will develop within the context of such a media recommendation and valuation environment. Note that initially 'anonymous' offerings are even possible, so that brand new mental associates might surface from time to time through a review of those who have recently offered the most astute suggestions.

However this filtration and sorting process is managed, the final media presentation would then be personalised into the form chosen as most useful and attractive to the end user. Of course commercially-based links could be seamlessly integrated into the same protocol. A standing subscription could retrieve the 'Top' stories from a trusted news source. Informative ads from a company of interest, or a billing for the latest film from a favourite director, might appear in a sidebar at the appropriate location within one's ëDaily Us'. And before too long we can expect to see well-known 'Mentors' selling their own personal selections of prioritised reference lists on various subjects, for automatic inclusion into the purchaser's online view of the world.

The scenario above would require some quorum of active participation to really take off, but the actual selection and interactive arrangement of the information could be carried out primarily by specifically designed software on each user's system. Only a minimum of new centralised service is required. The basic "Home" collection within the *HyperWave* environment for the Web might be adapted to support a first personal edition of ëThe Daily Us' [Maurer 1996]. Cooperative use of existing encryption standards like PGP could soothe any qualms that users might have about broadcasting 'sensitive' recommendations, but the encryption/decryption itself would add very little extra processing load since the reference links are so small. In the medium term, the maturation of Java could provide the basis for very rapid dissemination of software to implement such a system of interactive recommendation and presentation [Bank 1995].

Some resources on the Web already provide a 'thumbnail' image, along with a brief abstract about the content, and this will become commonplace as PDF continues to establish itself as a de facto standard. Having such a 'sampler' file about each online resource would provide a simple means with which to format the surface layer presentation of any 'accepted' suggestions. This image could either be sent as the recommendation itself, together with data on the referee of the link (perhaps along with their own specific annotation message), or be automatically downloaded only at the point of acceptance.

Valuation Process

One critical aspect of this overall scheme will be an intuitive means to provide feedback about the accuracy of any recommendations coming through the system, essentially to 'Vote' on the relevance of each media resource encountered. At first this would just be self-contained; giving the local assessment database a means to adapt its 'rating' of each referee in terms of their success at tagging appropriate resources, from the point of view of the recipient. Once there is some critical mass of participation in such a systematic sharing of value judgments, however, new possibilities begin to open up.

An 'impartial' survey organisation could regularly issue nearly every user's system with a set number of untraceable, but completely authenticated, units of 'judgment' (probably by way of offering participants some valuable incentive). Existing digital cash theory and initiatives provide conceptual models for this [Chaum 1992]. These non-negotiable units could be automatically 'stamped' by one's own system, based on the same criteria used for 'internal' assessment, then automatically returned to the issuing organisation. Negative value judgments would 'cost' the user in the same way, but be subtracted from the score of the online resource. The final, continuously updated results would clearly be of great importance to the commercial sector - and probably to human culture as a whole - by revealing the varied focus of collective human consciousness in very fine detail.

However the technical details evolve, large personal archives of these suggested links would naturally form, both of inward (enabling searches of 'back issues', even for material that was initially passed over) and also of outward suggestions (as a track on one's own media experience and personal networking). These personal archives would be "large" only in the sense of the number of entries, the digital storage space required for such hypermedia references is relatively modest. It will be nearly essential, however, for the overall environment to continue upgrading toward an interconnection protocol that uses bidirectional links - in order to avoid having many of those archived recommendations become useless because the suggested destination has moved or been updated in the interim [Fenn 1994].

Interface Design

All of the patterns of interaction described above will require an effective graphical interface with extensive potential for customisation by end-users. A critical function to optimise in the course of developing this is the means by which any particular media resource can be graded and recommended to any chosen set of recipients. The way in which one's recipient lists are managed is also crucial (and may eventually require a 3-D interface to adequately deal with the complexities of relationship that many people could generate).

One essential component of an interface that may come to provide a user's primary media access is adequate provision for ësocial priority' input. This would facilitate the throughput of material that has been narrowcast to anyone for whom it might be relevant, along with a suggested classification as to its priority. Any socially sound arrangement would still allow the individual user to selectively ëtemper' the precedence of different classes of such notification from different sources - even if this might not generally be seen as being in their own "best interests". The full ramifications of such questions about the right to personal isolation are beyond the scope of this paper, other than to note that they are likely to become a source of controversy in some societies.

In any case, the technical implementation of any such priority community-service access must be as reliable as possible; this being a feature which replaces one of the most critical functions of the mass media at present. For example, an emergency alarm that might contain notice about some local geographic event of interest - such as, "There's fire raging towards your house!" - would naturally be given a high priority by almost anyone (ëclass one'); whereas a reminder about some legal requirement such as filling in a census form might come through as medium priority (ëclass two'); and a general notice about rubbish collection could end up relegated to a background resource collection (ëclass three').

Initial Development Sequence

At the proof-of-concept and initial user-testing stage, an interface supporting the basic ërecommendation' functionality would be be constructed, along with a means for users to grade any media with which they're presented. Those taking part in this research could then intensively exchange recommendations with each other, being particularly diligent in searching out a wide range of material to suggest, since they'll be serving as the surrogate for future activity on the part of much larger and more diverse communities. At this first stage the processing of the recommendations themselves could be overseen by human 'agents', managing the functionality of an off-the-shelf database, and the final presentation layout might involve a graphic designer.

By far the most effective context in which to first implement an effective recommendation and valuation system for eventual use on the Web is probably within a large subscriber network. (Organisational intranets are generally too goal-directed and the wider Internet itself is too diffuse to readily develop a quorum of coordinated participation.) Within a context like a value-added IP service, however, development costs could potentially be offset through subscription charges; authentication of users is relatively simple; and all participants can readily be supplied with any required software updates.

Conclusion

It is imperative that we develop a cohesive conceptual framework for empowering users into self-structured contact with their own chosen online communities, while at the same time maintaining more generalised social cohesion. The separate trends leading us towards very widespread interchange of of resources and recommendations will continue to accelerate anyway, but without overall strategic guidance the result may prove in retrospect to be an undesirable or even dangerous patchwork. The proposed interactive structure offers flexibility in terms of commercial development and individual presentation, but maintains an integrated set of features to serve future social purposes. The authors suggest that intentional development of such an interface is the logical next step in achieving social efficacy and personal communications in a "post mass-media dependent "age. Looking further ahead, it is possible to discern the outlines of a future in which the natural evolution of this sort of ëreference' system will come to be used for interactively sharing streams of media derived from each individual's daily existence - giving us a chance to selectively ëtune in' to each other's experience of the world.

References

[Bank 1995] Bank, D. (December 1995). Java Saga. Wired 3.12 , 166-9 238-244

[Chaum 1992] Chaum, D. (August 1992). Achieving Electronic Privacy. Scientific American 272 , 96-101

[Fenn 1994] Fenn, B. (October 1994). Harmony on an Expanding Net, interactions 1.4, ACM Press, 29-38

[Maurer 1996] Maurer, H. (1996). HyperWave: A Second Generation Web Solution, Addison-Wesley, 31-118

[ëmoose' 1996] ëmoose' (1996). The Cancelmoose Home Page, http://www.cm.org/faq.html

[Negroponte 1995] Negroponte, N. (1995). being digital, Hodder & Stoughton, 152-156

3D Visualization of WWW Semantic Content for Browsing and Query Formulation

Richard H. Fowler, Wendy A. L. Fowler, and Jorge L. Williams
Department of Computer Science
University of Texas - Pan American
Edinburg, TX, USA 78539-2999
email: {fowler, wfowler, jwilliams}@panam.edu

Abstract: Visualization is a promising technique for both enhancing users' perception of structure in the Internet and providing navigation facilities for its large information spaces. This paper describes an application of the Document Explorer to the visualization of WWW content structure. The system provides visualization, browsing, and query formulation mechanisms based on documents' semantic content. These mechanisms complement text and link based search by supplying a visual search and query formulation environment using semantic associations among documents. The user can view and interact with visual representations of WWW document relations to traverse this derived document space. The relationships among individual keywords in the documents are also represented visually to support query formulation by direct manipulation of content words in the document set. A suite of tools is provided for navigation and orientation in the visual network representations of document set and term collection. ... [Additionally, Finally, the system provides direct manipulation facilities for forming queries to identify starting points in the networl of documents or use conventional vector-based retrieval mechanisms.]

Introduction

The explosive growth of the Internet has added to the need to filter and organize information so that users can efficiently and effectively identify relevant documents. Browsers such as Mosaic [Schatz & Hardin, 1994] for the World Wide Web and Harmony [Andrews & Kappe, 1994] for Hyper-G allow users to traverse the information space of documents through connections provided by document authors. Though this may account for much of the popularity of the Internet, it hides many of the difficulties of effective information access. Searching Internet resources is difficult due to size, diversity of data, and lack of a common indexing scheme. The most widely used search tools are automatically generated search systems like Lycos and Harvest and manually organized systems like Yahoo and Internet Yellow pages. Though such tools are valuable assets to the Internet searcher, it seems likely that these tools alone will not solve the current problems of information access.

The challenges of information access on the Internet are issues common to all forms of information retrieval. These issues include difficulties in using indexing vocabularies, indexing indeterminacy, and the user's inability to completely specify information needs [Ingwerson & Wormell, 1986]. Retrieving information that meets users' information needs is an iterative process, and techniques which explicitly incorporate users' judgments, such as relevance feedback [Maron & Kuhn, 1960], provide means to automate some aspects of user guided retrieval. It is also clear that mechanisms providing alternative paths of access to information can enhance retrieval effectiveness [Bates, 1986].

One promising approach for enhancing information retrieval through the Internet is visualization to facilitate users' perception of document relation structure. A number of systems have been developed to provide visually based browsing mechanisms for traversing the link structure of Internet documents. McCahill and Erickson [McCahill & Erickson, 1995] describe designs for three-dimensional spatial interfaces for Internet Gopher which employ icons with various shapes and textures to represent documents arranged by file structure or search results. The Harmony browser for Hyper-G [Andrews, 1995] provides two-dimensional structure maps of a document's link neighborhood, as well the three-dimensional Information Landscape visualization of link structure hierarchy. The Narcissus system [Hendley et al., 1995] creates a three-dimensional visual representation of WWW link connectivity with documents' positions determined by user manipulable attractive and repulsive forces among document nodes. Munzner and Burchard [Munzner & Burchard, 1996] describe a system which displays WWW documents positioned in a hyperbolic space to increase display density compared to conventional display spaces.

Visualizations of document connectivity based on existing, author-created links provide an important component in facilitating orientation and navigation in the large WWW information space. However, for users it is typically documents' content relationships which are of most interest, rather than link based relationships, and such semantic relationships are only partially reflected by link structure. Documents with similar content can be found throughout the WWW information space. Though it is likely that a user would find documents dealing with similar topics close together within a site, it is also likely that other documents dealing with similar topics are at other sites, and those related documents may be buried deep within the link structure of a site.

Systems which provide facilities for organizing and viewing document organizations based on semantic content are less common than systems which focus on visualizing document link structure. Mukherjea, Foley, and Hudson [Mukherjea et al., 1995] describe a system which operates on the semantic content of WWW documents to form visually displayed hierarchies. In this system users can specify attributes which affect the organization of documents derived by the system and reconfigure the display to suit their needs. VR-VIBE [Benford et al., 1995] is a system which fixes WWW documents at locations in three-dimensional space by allowing users to interactively position keywords on a pyramid. Document locations are determined by calculating the distance for each document from keywords' locations on the pyramid. In this way users create visually distinct clusters of documents for further inspection. LyberWorld's Relevance Sphere [Hemmje et al., 1994] also uses this general approach for forming three-dimensional document cluster displays. Gershon et al. [Gershon et al., 1995] describe a system which allows users to view an organization of documents visited as a hierarchy of links and also to construct a separate hierarchy based on their own needs. Additionally, the system creates a co-occurrence map of terms in documents visited to help users identify word patterns useful in a more fine-grained search.

This paper describes an application of the Document Explorer system to the visualization of WWW content structure. The system bases the visualization, browsing, and query formulation mechanisms on the semantic content of documents. These mechanisms complement text and link based search by supplying a visual search environment using semantic associations among WWW documents. The system operates on sets of WWW documents to construct a visually displayed network of documents based on content similarity. The user can view and interact with the visual representations to traverse this derived document space. The relationships among individual keywords in the documents are also represented visually to support query formulation by direct manipulation of content words from the document set. A suite of tools is provided which focuses on navigation and orientation in the visual representations of document set and term collection.

Visualizing WWW Semantic Content

Information visualization shares with scientific visualization the goal of providing understanding and insight using visual representations, yet is often faced with challenges not encountered in scientific visualization. For scientific visualization the visual representation is derived from a physical phenomenon, hence, there typically exists a natural visual representation based on spatial, temporal, or other properties of the phenomenon. Semantic relationships have no physical component which might naturally supply a basis for visualization [Gershon, 1994]. Rather, the elements to be visualized typically have only semantic properties with no inherently spatial analog from which to create a visual representation. The spatial ordering for semantic relations must be created as part of the information visualization process. From this perspective, visualization of semantic information spaces must provide: 1) spatialization of the abstract data, which may entail both data organization and derivation of a visual spatial representation of the data, 2) presentation of the spatial representation in a display space for user interaction and viewing, such as Euclidean two- or three-dimensional space, and 3) techniques and tools for user interaction with the visual representation.

The Document Explorer provides mechanisms for each of these three components of information visualization to display a semantic space of WWW documents and keywords. WWW document relationships and keyword relationships are organized as associative networks based on content similarity among documents and frequency of co-occurrence of keywords. The networks are presented in three-dimensional display space with network nodes positioned using a layout algorithm based on a mechanical spring metaphor. A variety of interaction and display tools are used to supply orientation cues and navigation mechanisms to assist users in browsing the document and keyword networks, as well as in the formulation of queries by direct manipulation of system objects.

Extracting and Organizing WWW Semantic Content

The system's principal visualizations are network displays based on documents' keyword lists. The lists can be provided by automatic content extraction tools, such as Harvest [Bowman et al., 1994], or derived from documents retrieved by the system. Keyword lists for each document are used to determine the associations among documents and among terms using co-occurrence metrics to derive similarity measures among documents and among keywords. Both visualizations of document space and term space for WWW document sets are available to the user.

The content based network of WWW documents is shown below in Figure 1 as the main window of the screen display. In this view documents are labeled by their content. Alternatively, the display can present the HTML title. Below are overview diagrams showing the location of the detailed view in the main window within the complete network. To the left and right of the overview is a series of visual bookmarks set while browsing which can be used to return to a previous viewpoint. The leftmost shows a view of the complete network. Other navigation and orientation tools are also available, such as anchors and signposts, which the user can leave and revisit at points traversed in the network. Change of viewpoint using navigation aids is always done by zooming to new viewpoints to maintain fluid motion and attenuate disorientation. In the upper right of the screen is a natural language query which has been transformed to a user-manipulable query graph. The query can be used to supply an entry point in the document network or be used for conventional weighted vector search to provide a list of documents from which to initiate browsing.

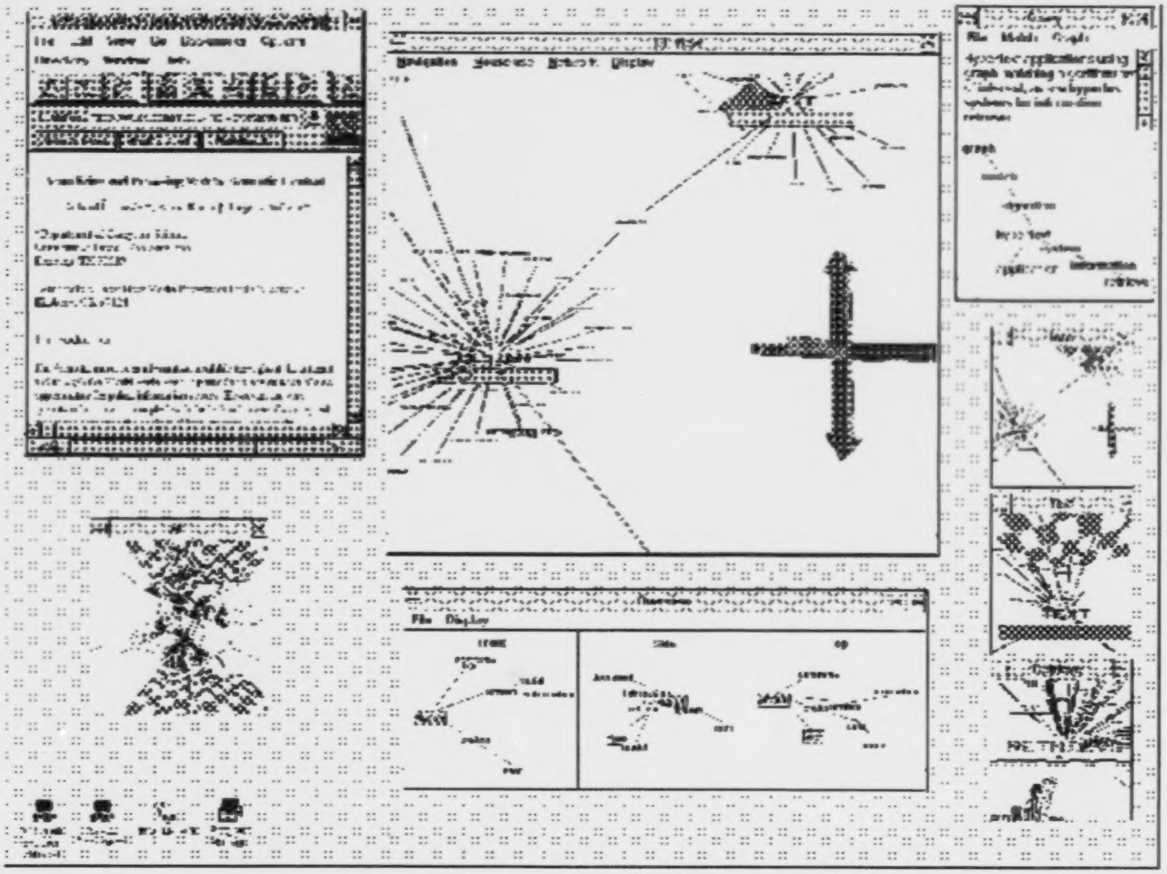


Figure 1: Document Explorer screen. The content based document network is shown in the main window with overview diagrams below it. Windows to the left and right of the overview show visual bookmarks the user has set. At the upper right is the user's natural language query and its visual representation.

The representations underlying the system's network displays are minimum cost networks derived from measures of term and document associations. The network of documents is based on interdocument similarity, as measured by co-occurrence of keywords between document pairs. For the network of terms, or associative term thesaurus, and the visual representation of the user's query the associations are derived from text with association measured by keyword co-occurrence and lexical distance within documents. The networks used in the system are Pathfinder networks (PfNets) [Dearholt & Schvaneveldt, 1990], which can be conceptualized as path length limited minimum cost networks. Algorithms to derive minimum cost spanning trees have only the constraints that the network is connected and cost, as measured by the sum of link weights, is a minimum. For PfNets, an additional constraint is added: Not only must the graph be connected and minimum cost, but also that the longest path length, as measured by number of links, is less than some criterion.

Reducing of the complexity of network representations is a central objective in many efforts in visualizing Internet information structures. Complexity can be reduced by transforming the network of connectivity among documents to a hierarchy by removing links. Compared to the more general problem of representing directed graphs, visualization of hierarchies is relatively well developed. This sort of transformation is used by Mukherjea et al. [Mukherjea et al., 1995] for WWW documents. In the Document Explorer nodes in the complete network are identified for display and navigation in a fashion similar to the identification of cluster centroids in single link clustering. These document nodes are continuously displayed in the overview diagram and the three-dimensional space within which the user navigates. The location of the subnetwork displayed in the main viewing window is updated

BEST COPY AVAILABLE

continuously in the overview diagrams as the user moves through the three-dimensional network display space.

Additionally, the system uses networks of differing densities to provide separate views of document and term interrelationships. The least dense PfNet, which allows paths of any length in satisfying the minimum cost criterion, is useful for global navigation and orientation. Conceptually, it shows the strongest relations among elements. As such, it is effective in supplying views for global navigation and structure perception. To provide a more detailed view of relationships the system also maintains a PfNet of elements in which the maximum path length is relatively small, creating a complementary network with many more links. This more dense network display is most beneficial when viewing a small set of elements and serves as a sort of magnifying glass for revealing relations among elements not shown in the sparse network.

Spatial representation of the system's networks

The spatial representation of the system's networks is designed to visually reveal structure. Network nodes are positioned in three dimensions using a graph layout algorithm based on a spring metaphor [Kumar & Fowler, 1994] similar to Kamada and Kawai's two-dimensional network layout algorithm [Kamada & Kawai, 1989]. Nodes are treated as connectors and spring length and strength among connectors is derived from network link distances. Nodes are allowed to vary in three dimensions and iteratively positioned at the points which minimize energy in the system of springs. Varying spring length and strength allows layouts which are useful for user interaction and visually reveal clustering and connectivity among.

Display and interaction mechanisms

The size and density of the system's WWW document network requires viewing and navigation tools that allow users to perceive the overall structure of document relations, explore smaller regions in detail, and select and view individual documents. Display and interaction mechanisms in the Document Explorer supply orientation and overview of the global structure of document associations, together with navigation and retrieval tools for exploring local detail. The overview diagram of the complete network is constructed using the nodes of highest degree in the network. As the user changes viewpoint in the main viewing window, and thus the portion of the network which is viewed in detail, the overview diagram tracks the overview nodes which are visible in the view volume of the detailed view. This helps attenuate disorientation by providing the context for the individual network nodes which are in view. We have also explored stereoscopic display of the networks. For the sorts of network displays in the Document Explorer there is a significant increase in the user's ability to perceive structure, as found in recent studies of stereoscopic network viewing [Ware & Franck, 1996].

Other display mechanisms for changing the user's views of the network are also designed to assist the user in remaining oriented to the overall structure while examining local detail, such as the availability of two separate networks which differ in density, or the number of links. The most sparse is essentially a tree, thus having the fewest links necessary to have all nodes connected. It is this sparse network which is used to provide the spring analogs used to position the nodes in three-dimensional space. In practice the arrangement of nodes on this basis supplies a characteristic clustering of nodes in which structure is relatively easy to perceive, such as shown in the leftmost bookmark of Figure 1. Yet, much of the utility of browsing in the derived semantic space comes from exploring relations which are relatively weak. These relations are captured in the more dense network. The user typically identifies a single document or document region by browsing and traversing links from a known document in the sparse network, and then selects to have the weaker links of a node or node set become visible. These are identified by a differing color and the point at which the selection is made is identified by a marker or signpost which is relatively large and can be easily zoomed back to during the course of exploration. Most simply, the user can select a node which represents a single document or term and expand or collapse connected nodes. Again, color is used to mark the course of browsing and network exploration.

Several navigational tools are available to the user which provide mechanisms for returning to previous viewing points, or supply information about the context of the current position. Visual bookmarks can be set by the user at any point while moving through the network. Simply selecting the bookmark returns the user to the viewpoint at that time the bookmark was set. As mentioned above, visual anchors can be set which remain visible from long viewing distances and can be returned to at later times. Other display mechanisms include link and node colorings indicating parts of the networks of high similarity to the user's query and user positioned navigational signposts pointing to the closest overview nodes or bookmarks. All navigation maintains fluid movement in the space, always zooming, rather than jumping, to new viewpoints.

Query Formulation

Although the system's most novel features center on its visual representations to support browsing and structure perception, it is useful for information retrieval systems to supply multiple paths of access to information items. For the type of minimum cost networks used in this system, there is a close relation to various clustering algorithms. Using the visual document space representation to browse documents can be characterized as a form of user directed, cluster based search. The system also supplies conventional vector space retrieval as an adjunct, supported by direct manipulation techniques for forming queries.

In a typical use of the system the user enters a natural language statement of information need to begin the browsing and retrieval process. The system converts the natural language to a weighted vector representation and uses conventional weighted vector search to form a sequence of documents matching a vector representation of the query. The user can then select a document from this list to serve as the entry point in the network of documents, i.e., the viewpoint is positioned near that document and the document is distinguished by color.

The availability of an association map of keyword relations, a term space, assists the user in formulating queries. The relations among terms as portrayed in an association map are typically quite different than found in a conventional thesaurus and serve as an alternative ordering of term relations. To support query formulation the system provides direct manipulation facilities for constructing queries. A visual, manipulable representation of the user's query is displayed together with the natural language from which it was originally derived. This visual representation can be manipulated by the user by moving terms from the term network into the query window and connecting them to the query graph. The vector space retrieval is based on the query graph and is performed by converting it to a weighted vector representation. This functionality could be provided by other interaction mechanisms, yet the visual graph manipulation technique encourages active interaction generally, and allows the of use network representations throughout the system.

Acknowledgments

Work on this project has been supported by NASA grants NAG9-551 and NAG9-842.

Conclusion

The Document Explorer supplies visualization, browsing, and query formulation mechanisms based on the semantic content of WWW documents. Users can view and interact with a visually displayed network of documents based on content similarity in a WWW information space that is an alternative to link based representations. Relationships among individual keywords in the documents are also displayed visually to support query formulation by direct manipulation and convey information about the keyword set. Navigation and orientation tools facilitate interaction and enhance perception of the document set and term collection structures through visual representations which facilitate the exploration of local detail while remaining oriented to the global context. These facilities can serve as useful additions to textual representations and visual representations based on document link structure.

References

[Andrews, 1995] Andrews, K. (1995). Visualising cyberspace: Information visualisation in the Harmony Internet browser. *Proceedings of Information Visualization*, 97-104. IEEE.

[Andrews & Kappe, 1994] Andrews, K. & Kappe, E. (1994). Soaring through hyperspace: A snapshot of Hyper-G and its Harmony client. *Proceedings of Eurographics Symposium of Multimedia/Hypermedia in Open Distributed Environments*, 181-191. Springer Verlag.

[Bates, 1986] Bates, M. J. (1986). Subject access in online catalogs: A design model. *Journal of the American Society for Information Science, 37(6),* 357-386.

[Benford et al., 1995] Benford, S., Snowden, S., Greenhalgh, C., Ingram, R., Knox, I., & Brown, C. (1995). VR-VIBE: A virtual environment for co-operative information retrieval. *Eurographics '95,* 349-360.

[Bowman et al., 1994] Bowman, C. M. et al. (1994). The Harvest information and discovery system. *Proceedings of the Second International WWW Conference, WWW '94*, 763-771.

[Dearholt & Schvaneveldt, 1990] Dearholt, D. W. & Schvaneveldt, R. W. (1990). Properties of Pathfinder networks. In R. W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization*, 1-30. Ablex.

[Gershon, 1994] Gershon, N. D. (1995). (Panel chair.) Information visualization: The next frontier. *Proceedings of SIGGRAPH '94,* 485-486. ACM.

[Gershon et al., 1995] Gershon, N. D., LeVesseur, J., Winstead, J., Croall, J., Pernick, A., & Ruh, W. (1995). Visualizing Internet resources. *Proceedings of Information Visualization*, 122-128. IEEE.

[Hemmje et al., 1994] Hemmje, M., Kunkel, C., & Willett, A. (1994). LyberWorld - A visualization user interface supporting fulltext retrieval. *Proceedings of ACM SIGIR*, 249-259. ACM.

[Hendley et al., 1995] Hendley, R. J., Drew, N. S., Wood, A. M., & Beale, R. (1995). Narcissus: Visualizing information. *Proceedings of Information Visualization*, 90-97. IEEE.

[Ingwerson & Wormell, 1986] Ingwerson, P., & Wormell, I. (1986). Improved subject access, browsing and scanning mechanisms in modern online ir. *Proceedings of ACM SIGIR*, 68-76. ACM.

[Kamada & Kawai, 1989] Kamada, T. & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information processing letters, 31*, 7-15.

[Kumar & Fowler, 1994] Kumar, A. & Fowler, R. H. (1994). *A spring modeling algorithm to position nodes of an undirected graph in three dimensions*, NAG9-551-4, Department of Computer Science, University of Texas - Pan American, Edinburg. Available at http://www.cs.panam.edu/info_vis.

[McCahill & Erickson, 1995] McCahill, M. P. & Erickson, T. (1995). Design for a 3D spatial user interface for Internet Gopher. *Proceedings of ED-MEDIA 95*, 39-44. AACE.

[Maron & Kuhn, 1960] Maron, M. E. & Kuhn, J. L. (1960). On relevance, probabilistic indexing, and information retrieval. *Journal of the Association for Computing Machinery, 7(3),* 216-244.

[Mukherjea et al., 1995] Mukherjea, S. Foley, J. D., & Hudson, S. (1995). Visualizing complex hypermedia networks through multiple hierarchical views. *Proceedings of CHI '95,* 331-337. ACM.

[Munzner & Burchard, 1996] Munzner, T. & Burchard, P. (1996). Visualizing the structure of the World Wide Web in 3D hyperbolic space. Available at http://www.geom.umn.edu/docs/weboogl/, The Geometry Center, University of Minnesota.

[Schatz & Hardin, 1994] Schatz, B. R. & Hardin, J, B. (1994). NCSA Mosaic and the World-Wide Web: Global hypermedia protocols for the Internet, *Science, 265*, 895-901.

[Ware & Franck, 1996] Ware, C. & Franck, G. (1996). Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics, 15(2),* 121-140.

Interactive Justice Through the Web

Giovanni Fulantelli, Mario Allegra
Istituto di Tecnologie Didattiche e Formative - Italian National Research Council
e-mail: {fulantelli, allegra}@itdfmac.itdf.pa.cnr.it

Abstract: In 1995, at the Italian National Research Council, we started to put on-line, through the World Wide Web, the original legal documentation of two important trials regarding the Sicily's Mafia and its connection with the political management. The main aims of the two developed web sites were to allow user to follow the evolution of the two trials through the complete acts of the Court and to set up a learning environment to deepen the Italian penal code, by analysing two real cases. Web users around the World have accessed and used the sites as on-line anti-Mafia movement, by reading the complete acts of the trials and directly publishing on the web their messages of solidarity to the victims' relatives and to the judges who are fighting against the Mafia.

Introduction

Information and communication technologies progresses are quickly leading towards the so called Information Society. Important social and cultural changes are occurring during this process, and must be taken in account carefully. Theoretical frameworks for these changes can be found in literature. Marshall McLuhan, introduced the fundamental idea of Global Village, by considering that the globe is no more than a village as electrically contracted.

Some years later, the role of the networks as shapers of social communities was clearly discussed by J.C.R. Licklider, one of the profets of the Net, that focused on the principles for computer networks to become communication tools. Nowadays, it is amazing/astonishing to note how the same principles are behind the Internet success.

In particular, the impact of the World Wide Web on the society can be very strong: through Internet, people can access information everywhere in the World and can communicate with everyone around the World; furthermore, the Web makes the Internet much more accessible to people, so that everyone can join the growing Internet community. According to McLuhan, the Web is contracting the time and space distances. Therefore, it is straightforward to consider this community as an important step towards the Global Village.

As a matter of fact, people information needs become greater and greater and networks permit users to access huge amounts of information in a simple and effective way; in this new community, citizens are much more active in selecting the information they are interested in; consequently, information providers have to create and develop more and more information-on-demand services in order to satisfy the users needs.

In addition, the more people use computer networks to access information sites, the more they feel the need to discuss their problems and questions with other citizens of the virtual networked community, and to be active player in the social, political and cultural life. According to their specific needs, they wish to talk to the Administration, rather than to experts or common citizens like themselves to exchange opinions with. The Civic Networks, information systems in Internet concerning the Administration, represent the most significant outcome of these needs. They allow citizens to get into civic and local problems, and to socialise with other people by talking about needs, problems, solutions, common initiatives.

Another important social sector asking for complete information is the Justice; international well-known facts and world-wide notorious mysteries require efficient information media which can be provided by computer networks.

Particularly, people accessing the original proceedings of high-profile trials through the networks can follow the trials in progress, read the public prosecutors and defence lawyers acts, and eventually get an idea by themselves rather than by the newspapers, which report filtered stories. Of course, the newspapers role in informing people is strongly important, as long as people can easily compare different points of view with the original trial facts. Consequently, it is extremely important to integrate original proceedings and extracts from the newspapers in the same documentation system.

The previous considerations convinced us to set up two Web sites on two trials against the Mafia.

The aim of this paper is to describe them and to highlight the social aspects of our initiative. In addition, we have designed a Distance Education course based on the two sites for University Law students, and we have set up a virtual classroom with students across the country. Finally, design principles of this course are reported.

The Trial Web Sites

During the past two years, two web sites regarding two extremely dramatic Italian trials have been set up at the Istituto di Tecnologie Didattiche e Formative in Palermo.

The first Web site was set up in April 1995, on the occasion of the beginning of the trial of some Sicily's Mafia members accused of murdering Giovanni Falcone, a crusading anti-Mafia judge, along with his wife and three bodyguards.

The web site, which has been named the "Praesidium Falcone", was set up in order to allow people all around the World, and mainly in Italy, to follow the trial in its evolution through original acts and different opinions. To this end, we got permission by the trial Public Prosecutors to publish in Internet the original and complete documentation. In addition, we included some articles from national and local newspapers in the Web site.

However, the World Wide Web is not simply an attractive publishing space; in fact, it has proved to be an extremely effective tool to implement interactive information systems, thus satisfying the more and more pressing needs of people to play an active role in public life. For this reason, we decided to allow users to send messages of solidarity to the victims' relatives and to the judges who are fighting against the Mafia; these messages are automatically published in a specific area of the site.

The second web site concerns the People vs. Giulio Andreotti trial. The trial of the country's seven-time Prime Minister, charged with conspiring with Sicily's Mafia, is a very important trial, because for the first time in Italy, Public Prosecutors publicly denounces a strict connection between the Sicily's Mafia and the Italian political management..

Because of the fame of the Italian politician, the Public Prosecutors' thesis has shared the popular opinion in pro and against the former Prime Minister. Consequently, an interactive section in the Web, like the one in the Praesidium Falcone , would have been easily transformed in an exploited section, with many and useless 'flames'. On the contrary, the Capaci trial is against well-known criminals (people already condemned for other murders and criminal affairs); consequently, the popular opinion is unanimous against the accused, and it seemed sensible to create an interactive section to allow users to report their feelings about the murder and the trial.

Yet, in order to let users participate somehow at the Andreotti trial and to smooth the dramatic power of the relative site, we preferred to activate an interactive section for this site where we ask only for satirical illustrations of the former Prime Minister.

A second important difference between the two Web sites concerns the documentation in them. As far as the Praesidium Falcone is concerned, we obtained the murder reconstruction and all the acts presented by the Public Prosecutors when the trial begun. As regards the Andreotti trial, we got formal authorisation to post the official proceedings on the Net by the Court; in this case, it is extremely significant to publish information presented by both public prosecutors and defence lawyers, thus avoiding a partial information and because several technical aspects are comprehensible only by analysing the two opposing interpretations.

Social Aspects of the Web Sites

Social aspects of the Web have being widely discussed over the past few years. To the aim of this paper, we are interested in the view of the Web as a natural place for a number of communities that, by adopting the Net as their virtual habit, communicate on specific topics, and which are tied by common interests. For example, for many Italians living abroad, the Net has become the source for news about Italy.

As far as the two web sites described in this paper, Web users around the World have accessed and used them as on-line anti-Mafia movement, by reading the complete documentation of the trials and sending messages of solidarity to the victims' relatives and to the judges who are fighting against the Mafia.

Since some sections of the sites are in Italian, French, English and German, messages from all around the World arrived; some of them were directly published in the Web, others were sent to the webmaster or some of us.

Many Italians living abroad and elsewhere in Italy have been especially outspoken on the Internet.

"I am a Sicilian student in Wales...and I came across this very beautiful initiative to remember the Judge Falcone.....Thanks to Falcone and Borsellino (another anti-Mafia judge murdered by the Mafia) and all people that lost the life in the fighting against the mafia." a Sicilian wrote.

A person with the surname as the murdered judgeí, while no relation to him, wrote: iI will not forget the man who made me proud over my name. My grandfather left Sicily 65 years ago. He died as a soldier in the Great War just as Giovanni Falcone. I wish I have had the opportunity to meet them both".

And many messages arrived from the rest of the World: an American wrote: "I wish 100 people were born for each Falcone or Borsellino murdered by the mafia."

Yet, the message from a mother: "We don't want to forget it, we mustn't forget it, all we have to devote ourselves to make our children remember and understand ......I am wife and mother. Women that have been wounded by the mafia, you have to know that you are not alone."

Particularly impressive the message of solidarity from a Belgian judge: "I am solidly behind the Italian judges".

Finally, the site's dual social purpose - to provide information and an emotional outlet - leaded us to bring a computer to the Piazza Magione in Palermo, where a demonstration was held in the dead judge's honour, on the third anniversary of the Falcone murder. Over a live Internet line, we received and read aloud messages arrived from the World. In addition, we put some images of the demonstration on-line, so that people in Internet could be virtually present to the demonstration, thus virtually extending the physical square boundaries to the whole Internet community. A similar solution was adopted on the occasion of the fourth anniversary of the Falcone murder.

The Structure of the Learning Section

Over the past months, a variety of courseware has appeared on the WWW; many courses provide lecture material in hypertext form and some of them include links to other documents world-wide. However, most of the Internet based courseware provides individual learning environments, without exploiting one the most effective educational feature provided by Internet: the interpersonal communication, basic to the co-operative learning model.

The aim of the learning section in the Web sites described in this paper is to allow students to deepen the Italian penal code, by analysing technical aspects of the two trials. Because of the availability of the acts of two trials through the Web, it has been straightforward to adopt the Web as the interaction environment for the actors of the didactic process.

But the choice of the web is due to other important reasons too; in fact, it allows for very easy interactions based on its multimedia interface and it allows the integration of the different Internet communication tools (e-mail, news groups, mailing-lists, and so on) in the same interaction environment. Consequently, the user can use an unique and simple tool to make the most of the different communication modes as established in the learning environment general model.

This choice leads to several interesting advantages both for teachers and students. The teacher can integrate, through the WWW, part of bibliographic material (whether directly input through the keyboard or through scanner and OCR, depending on the size and type of the documents) in his explanations; in this way, the teacher can build up a digital library structured according to his own needs and criteria, and which provides the basic elements to organise guided tours through the documentation; these tours can be easily adapted, time by time, to specific didactic contexts.

For example, in the presented learning environment, the teacher could include some extracts from the penal code, in order to explain and comment on Court decisions; similarly, he could enclose related decisions from different trials.

In turn, students must think about interpretations of the Italian penal code, about the technical arguments by the accuse and the defence, and about the final decisions taken by the Court.

Note that very effective strategies for learning can be easily implemented on the opposite interpretations by public prosecutors and defence lawyers.

The learning environment is based on real in progress cases; consequently, there are a lot of interesting aspects to be considered both at the design phase and during the learning process:

- the course content is only partially known at the beginning of the course and it will cover only some specific topics of the penal code;

- production and publishing of new sections of the course depend on the trial evolution, so that it is not always possible to know when they will be available;

- lessons and lectures are discussions on specific technical aspects of the trial;

- discussions and comments on the trial are stimulated not only by the teacher (for example, by proposing juridical interpretations of the trial or by stimulating the students to provide their interpretations), but also by the students (for example, by asking for opinions or clarification on the trial);

- it is possible to analyse technical aspects on which the Court has not yet expressed itself, and to compare the results of these analysis with the Court final. This approach has been proved to be very stimulating for students.

All these issues affect the organization of the course; teachers cannot prepare lessons at the beginning of the course because they have to wait for the trial acts as starting point for their lessons; they have to prepare some extra-lessons to keep high the attention of the during the pauses of the trial; they have to moderate discussions among students, guiding them towards the most technical matters of the trial.

On-line courses based on real cases will become more and more effective when there will be plenty of examples available on the Net, and teachers will be able to refer to them in order to analyse different facets of complex problems. In our case, when a lot of trials will be available on line, the teacher will refer to several interpretations of the same articles of the penal code, or to deepen the analysis of a juridical problem.

Letís analyse the organisation of the learning section of the Andreotti web site; it allows for various communication modes according to the general flexibility of CMC-based learning environment.

The first communication mode permits a teacher-student interaction based on a list of topics proposed by the teacher. To each topic corresponds a web page and every message sent (from the teacher or the students) regarding that topic, is added to that page. By means of this list, the teacher highlights

the most interesting technical aspects of the trials; students are encouraged to work by themselves or in group towards an answer or simply to discuss the topic. They can talk to everyone in the virtual classroom by sending messages through the web; their messages are automatically and immediately published in a web page reserved to that topic.

A different communication mode allows users to send messages to the lists directly from a web page, so that they can use the same interface for all their didactic activities; a list-server forwards messages to all the members of the list.

The learning environment is based on asynchronous communication modes and therefore one of the main advantages is the opportunity to involve people who would normally be too busy to undertake traditional classroom lessons; in our case we have involved experts from major Italian Tribunals to serve as on-line guest lecturers in the classroom. Students from several Italian Law Universities can easily have access to the trial documentation, actively participate in the sessions and join public and private e-mail discussions on topics generated during the lectures.

During the first months of this experience, we have limited the number of students to a small group and we have focused our attention on the communication environment, adding some functions to simplify the use of the web and analysing problems and interaction modes among students and teacher. For the next academic year we are planning to associate this course with the traditional one, at the University of Palermo. This will allow to better evaluate the effectiveness of this new approach.

Conclusions

Through the two trials on the Net, we have given, for the first time, the possibility to follow a trial during its evolution, through the analysis of the original legal documentation, newspapers articles and expertsí explanations. Indeed, the Net has become the source for news about high-profile Mafia trials in Italy. Thousands from around the World had log on, downloading on their PCs everything from the opening arguments to the latest comments by public prosecutors to stories by even the smallest of Sicilian newspapers.

We have reached some important results; firstly, the Web technology has allowed to spread news about Sicily's facts all around the World, in such a way to create a very strong anti-mafia movement; secondly, we allowed Italian people leaving abroad to be a bit closer to their country, by breaking the space barriers; finally, we allowed people to debate on public matters in a very efficient way.

In addition, we have developed a CMC-based course on a real trial in progress, by using the original proceedings of the trials available on two Web sites set up at our Institute as lecture material and exploiting the communication advantages of the Web.

The use of the World Wide Web as a tool to implement the educational environment has proved to be effective; it allows to create new opportunities for teaching, in particular for learning environment based on real in progress cases that help students to complete their learning process.

References

[Andrews et al, 95]. Andrews, K., Kappe, F., Maurer, H., Schmaranz, K. (1995). On Second Generation Network Hypermedia Systems. Proc. of ED-MEDIA 95, Graz, AUSTRIA.

[Berners-Lee et al, 94]. Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H.F. & Secret, A., (1994). The World-Wide Web. CACM, 37 (8).

[Bilotta et al, 95]. Bilotta, E., Fiorito, M., Iovane, D., Pantano, P. (1995). An Educational Environment Using WWW. Proceedings of the Third International Conference on WWW, 95, Damstaud, Germania, pub. Elsevier Science.

[Frivold 1995]. Frivold, T.J. (1995). Extending WWW for Synchronous Collaboration. SRI International

[Hughes 1994]. Hughes, K. (1994). Entering the World-Wide Web: A Guide to Cyberspace. World Wide Web document, URL: http://www.hcc.hawaii.edu/guide/www.guide.html

[Licklider and Taylor, 1968]. Licklider, J.C.R. & Taylor, R. (1968). The Computer as a Communication Device. Science and Technology, April 1968

[McLuhan 1964]. McLuhan M. (1964), Understanding Media: the extensions of man. London: Routledge.

[Nielsen 1995]. Nielsen, J. (1995). Multimedia and Hypertext: The Internet and beyond, Academic Press, San Diego, CA

[Paulse 1995]. Paulsen, M.F. (1995). An Overview of CMC and the Online Classroom in Distance Education, in Computer Mediated Communication and the Online Classroom - Vol. 3, Hampton Press, Inc.

[Rheingold 1993]. Reinold, H. (1983). The Virtual Community: Homesteading on the Electronic Frontier. Addison-Wesley.

[Strassel 1996]. Strassel, K.A. (1996). Some Avid Italians Watch Andreotti Get His Day in Cybercourt. The Wall Street Journal Europe, 16-17/2/1996

Acknowledgements

We would like to thank our colleagues Onofrio Di Giuseppe and Marco Arrigo, who have implemented the two web sites and the virtual classroom, for their important comments on the whole project. We would like also to thank Prof. Antonio Greco for his support and contributions and Prof. Alfredo Galasso for his important remarks on the design of the learning section. Finally, we would like to thank all the people that contributed to the organization of the telematic praesidiums during the demonstrations against the Mafia.

Convergence to the Information Highway

Brian R. Gaines

Knowledge Science Institute
University of Calgary
Alberta, Canada T2N 1N4
gaines@cpsc.ucalgary.ca
http://ksi.cpsc.ucalgary.ca/KSI/

Abstract: The convergence of telecommunications and computing technologies/ services to a new medium offering integrated services through digital networks was predicted in the 1970s and is beginning to have major social and commercial impacts in the 1990s. This article analyzes the technological infrastructure of convergence to an information highway, tracing the origins of the concept, the false starts, the origins and growth of the Internet and World Wide Web, convergence as a substitution process, and the learning curves of the technologies involved.

- 1 The Path to the Information Highway
- 2 The Growth of the Internet and World-Wide Web
- 3 Convergence as a Technology Substitution Process
- 4 The Learning Infrastructure of Information Technology
- 5 Conclusions
- References

1 The Path to the Information Highway

The motivation for an "information highway" was expressed in 1937, just prior to the advent of computer technology, when Wells was promoting the concept of a "world brain" based on a "permanent world encyclopaedia" as a social good through giving universal access to all of human knowledge. He remarks:

> *"our contemporary encyclopaedias are still in the coach-and-horses phase of development, rather than in the phase of the automobile and the aeroplane. Encyclopaedic enterprise has not kept pace with material progress. These observers realize that the modern facilities of transport, radio, photographic reproduction and so forth are rendering practicable a much more fully succinct and accessible assembly of facts and ideas than was ever possible before." (Wells, 1938)*

Bush, a technical advisor to Roosevelt, published in 1945 an article in *Atlantic Monthly* which highlighted problems in the growth of knowledge, and proposed a technological solution based on his concept of *memex*, a multimedia personal computer:

> *"Professionally, our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose...The difficulty seems to be not so much that we publish unduly in view of the extent and variety of present-day interests, but rather that publication has been extended far beyond our present ability to make real use of the record." (Bush, 1945)*

The world brain has continued for over fifty years to provide an active objective for the information systems community (Goodman, 1987), and memex is often quoted as having been realized fifty years later through the World Wide Web (Berners-Lee, Cailliau, Luotonen, Nielsen and Secret, 1994).

The advent of time-shared *conversational computing* (Gruenberger, 1967; Orr, 1968) in the early 1960s allowed computers to be used to begin to address these early visions in providing a *national information system* (Rubinoff, 1965) or a *computer utility* (Parkhill, 1966). Martin's model of a "wired society" in

1978 comes closest to forecasting many aspects and impacts of the information highway as it is envisioned today:

> *"In the past, communications networks have been built for telephony, telegraphy, and broadcasting. Now the technology of communications is changing in ways which will have impact on the entire fabric of society in both developed and developing nations. In the USA the technology revolution coincides with a change in the political and legal structure of the telecommunications industry; the combination is explosive. Some countries will take advantage of the new technology; some will not. Some businessmen will make fortunes. Some companies will be bankrupted." (Martin, 1978)*

However, attempts to make available the wired society at the time of Martin's seminal work were presented in terms of greatly exaggerated expectations. For example, in 1979 Fedida and Malik present Viewdata as having the potential to have major social and economic impacts:

> *"We believe that Viewdata is a major new medium according to the McLuhan definition; one comparable with print, radio, and television, and which could have as significant effects on society and our lives as those did and still do. Like them, it may well lead to major changes in social habits and styles of life, and have long-lasting as well as complex economic effects." (Fedida and Malik, 1979)*

Other books of the same period describe the commercial, social and educational potential of Viewdata and interactive Videotex in similar glowing terms (Sigel, 1980; Woolfe, 1980; Chorafas, 1981; Winsbury, 1981), but the potential never materialized although systems such as Minitel in France may be seen as primitive ancestors of the information highway.

It was not until the 1990s and the advent of the World Wide Web that a system with many of the attributes of Well's world brain and Bush's memex came into being. The web makes available linked and indexed interactive multimedia documents so that it emulates the printed publication medium but also goes beyond it in offering sound, video and interactivity. Its founders describe it in terms reminiscent of Wells' vision:

> *"The World Wide Web (W3) was developed to be a pool of human knowledge, which would allow collaborators in remote sites to share their ideas and all aspects of a common project" (Berners-Lee et al., 1994)*

However, even an active and interactive encyclopaedia is an inadequate model of existing Internet activities because it neglects the integration of human-to-human discourse. Newsgroups and list servers support for mutual interest communities where questions are answered not be consulting an encyclopaedia but by consulting other people. This corresponds to another prophesy from the early days of timeshared computing:

> *"No company offering time-shared computer services has yet taken advantage of the communion possible between all users of the machine...If fifty percent of the world's population are connected through terminals, then questions from one location may be answered not by access to an internal data-base but by routing them to users elsewhere--who better to answer a question on abstruse Chinese history than an abstruse Chinese historian." (Gaines, 1971)*

Wells and Bush described the implementation of their visions in terms of the media technologies of their time and did not foresee the advent of television and its impact as a source of knowledge (Bianculli, 1992). They also neglect human discourse as another significant source of knowledge, human society as a *living encyclopaedia*. The envisioned information highway may be seen as an extended "world brain" accessed though the personal computer as a "memex" that integrates all available media and means of discourse to give active presentations of, and interactive access to, all of human knowledge. The current facilities of the Internet and World Wide Web provide a primitive implementation of the highway.

2 The Growth of the Internet and World-Wide Web

The Internet and the World Wide Web both typify technologies that come into being through serendipity rather than design in that the intentions and aspirations of their originators had little relation to what they have become. As the development of the electronic digital computer can be attributed to needs and funding in the 1940s arising out of the second world war, so can that of the Internet be attributed to needs and funding in the 1960s arising out of the cold war. The Eisenhower administration reacted to the USSR launch of Sputnik in 1957 with the formation of the Advanced Research Projects Agency to regain a lead in advanced technology. In 1969 ARPANET (Salus, 1995) was commissioned for research into networking. By 1971 it had 15 nodes connecting 23 computers and by 1973 international connections to the UK and Norway. In 1984 the National Science Foundation funded the creation of a national academic infrastructure connecting university computers in a network, NSFNET. In 1987 the net had grown to such an extent that NSF subcontracted its operation to Merit, and in 1993/1994 the network was privatized.

The number of computers connected through the Internet has grown from some 28 thousand at the beginning of 1988, to over 9 million at the beginning of 1996. Figure 1 shows data plotted from the Internet Domain Surveys undertaken by Network Wizards (NW, 1996).

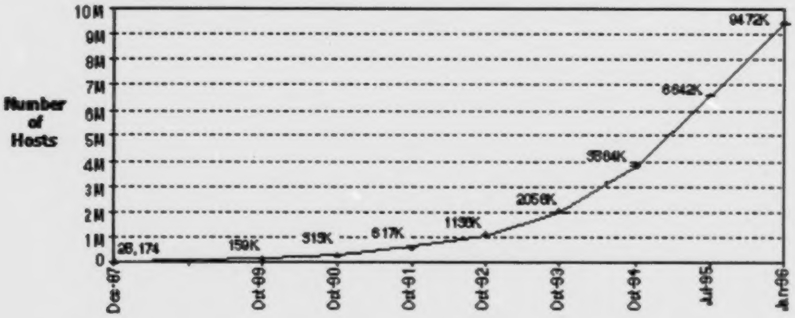


Figure 1 Growth in number of hosts on the Internet

The World Wide Web was conceived by Berners-Lee in March 1989 (CERN, 1994) as a "hypertext project" to organize documents at CERN in an information retrieval system (Berners-Lee and Cailliau, 1990). The design involved: a simple hypertext markup language that authors could enter through a word processor; distributed servers running on machines anywhere on the network; and access through any terminal, even line mode browsers. The web today still conforms to this basic model. A poster and demonstration at the ACM Hypertext conference in December 1991 announced the web to the computing community. However, major usage only began to grow with the February 1993 release of Andreessen's Mosaic for X-Windows. Whereas the original web proposal specifically states it will not aim to "do research into fancy multimedia facilities such as sound and video" (Berners-Lee and Cailliau, 1990), the HTTP protocol for document transmission was designed to be content neutral and as well-suited to multimedia material as to text.

In March 1993 the web was still being presented (Berners-Lee, 1993) as primarily a hypermedia retrieval system, but in November that year a development took place that so changed the nature of the web as to constitute a major new invention in its own right. Andreessen issued NCSA Mosaic 2 using tags to encode definitions of Motif widgets embedded within a hypermedia document, and allowed the state of those widgets within the client to be transmitted to the server. Suddenly the web protocols transcended their original conception to become the basis of general interactive, distributed, client-server information systems. This change was again serendipitous since the original objective of the design had been to enable the user to specify retrieval information in a dialog box that was embedded in a document rather than in a separate window. The capability of the user to use a web document to communicate with the server is the basis of commercial transaction processing on the web.

The growth rate of overall Internet traffic is some 100% a year. However, web traffic was growing when last accurately measurable in 1994 at some 1,000% a year. The growth of the web relative to all the other services is apparent if one plots the proportion of the data accounted for by each service. Figure 2 shows the proportion of FTP, web, Gopher, News, Mail, Telnet, IRC and DNS traffic on the NSFNET backbone from December 1992 through April 1995. It can be seen that the proportion of all services except FTP and HTTP remain relatively constant throughout the period, declining slightly towards the end. However, the proportion attributable to FTP decreases while that due to the web HTTP protocol increases and becomes greater than that through: IRC in October 1993; Gopher in March 1994; mail in July 1994; news in November 1994; and FTP in March 1995. This corresponds to the basic web protocol becoming the primary carrier of net data traffic with a 25% and growing share when last measurable.

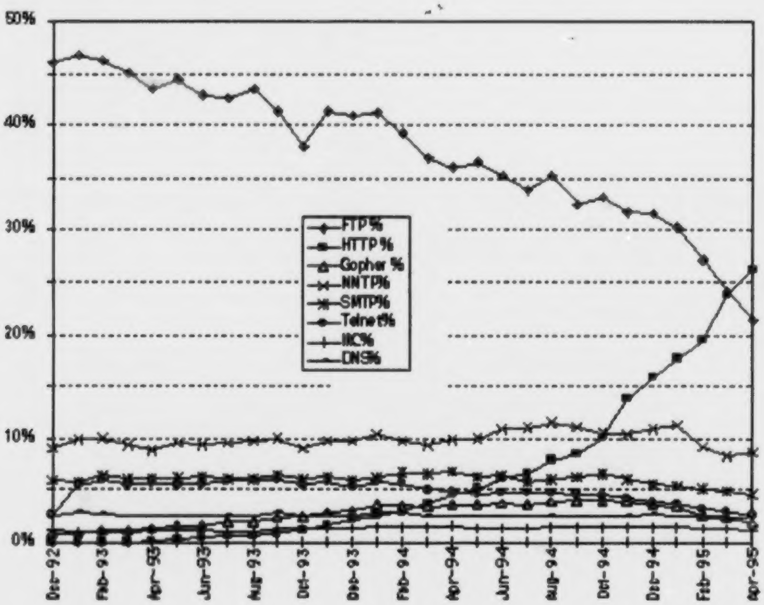


Figure 2 Proportion of FTP, web (HTTP), Gopher, News (NNTP), Mail (SMTP), Telnet, IRC and DNS traffic on the NSFNET backbone 1992-1995

Experience to date makes it problematic to characterize the commercial potential of the information highway, its actual social impact, and whether the protocols, technologies, carriers and equipment on which the present implementation is based are an adequate basis for future development. The information systems industry is well-known for over-ambitious expectations of technologies and their impact--a decade ago `expert systems' were going to revolutionize industry and create a new five billion dollar industry--they did neither. The `video phone' and `voice typewriter' have been on the horizon for over thirty years but have found no market and no adequate technology, respectively. It is eminently reasonable to be suspicious of claims that the information highway will be the driver of the next economy, and that the technology to implement it is practically in place.

The remaining sections present the substitution and learning processes of the information technologies underlying the information highway in order to provide a basis for forecasting the time scales of convergence and its socio-economic impact.

3 Convergence as a Technology Substitution Process

Telecommunications and computing technologies have common roots in electronics device technology, exploiting it to provide systems and services in similar ways. The early histories of the typewriter, the phonograph, the telephone, the movie, the computer and radio and television broadcasting share much the same timelines from the late nineteenth through twentieth century. This common background is

BEST COPY AVAILABLE

based on a technological progression from mechanical devices through vacuum tubes to the transistor and the integrated circuit containing a number of transistors (Braun and Macdonald, 1978). The transition from mechanical to electronic motion leads to increased reliability and greatly increased speed of operation. However, all of the applications above have also been involved in three other transitions of equal significance. The first is the transition from analog electronic signals directly representing the continuous variables involved to a digital electronic signal *encoding* those variables. The second is the transition from special-purpose computing architectures where the circuits are designed for a particular task, to general-purpose computing architectures which can be *programmed* for a particular task. The third is the transition from programs as fixed circuits to programs as themselves variable digital data allowing general-purpose machines to be simply *reprogrammed* for different tasks.

Whereas the transition from mechanical to electronic devices may be seen as a simple substitution of a faster more reliable technology, and that from analog to digital may be seen as a simple substitution of a more precise and reliable technology, the profound impact of transition to programming was serendipitous because the basic reason for doing it was to improve reliability by using fewer electronic components. In the early 1940s the concept emerged that the use of unreliable electronic components could be minimized by using a sequence of instructions to carry out a complex operation (Mauchly, 1942).

Programming gives to computers a fundamental property of being, in some sense, universal machines, since they can be programmed to emulate the calculation performed by any other machine. This means that, in principle, any special-purpose electronic circuit can be substituted by a general-purpose computer together with a program to emulate the operation of the special purpose circuit. For the electronics industry this makes possible economies of scale through very high-volume mass production of computer chips that can be functionally specialized to a wide range of applications as required. For customers it makes it possible to purchase one general-purpose system that can be used for a variety of functions, some of which may be unknown at the time of purchase.

The economic logic for substituting special-purpose systems with general-purpose programmed systems is straightforward. At a given state of the art in circuit technology the counter-balancing considerations are:-

1. Negatively, the programmed system is generally slower (by a factor of between 10 and 100) than the special-purpose system-- a potential decrease in performance.
2. Negatively, the programmed system generally uses more storage than the special-purpose system-- a potential increase in cost.
3. Positively, the programmed system is more widely applicable and subject to mass production-- a potential decrease in cost.
4. Positively, the programmed system is multi-functional-- a potential decrease in cost and improvement in capabilities.

Consideration 1 is a fundamental limitation since, if the application requires processing speeds not feasible with existing computer technology, then substitution cannot occur. This is why digital television is a late arrival since the data rates required for television signal processing have been too high for mass market computer products until the mid-1990s. Consideration 2 is also a limitation in television applications such as digital editors where the cost of random-access digital storage is still high compared with analog tape. However, as technology improves, in applications to which these limitations come not to apply the logics of consideration 3, mass production, and 4, versatility, lead to inexorable substitution.

The diffusion of computer applications may be seen as a process of substitution of electronics for mechanics, followed by a process of substitution of general-purpose programmed electronics for special-purpose electronics. Special-purpose circuits remain essential for applications where the rate of information processing exceeds that possible in a low-cost general-purpose computer, or where the cost of transforming the signals involved to and from digital form exceeds the cost savings of using a general-purpose computer. Convergence may be seen as a phenomenon of such substitution taking place in the consumer telecommunication markets with telephones, radio, television, VCR's, cameras, and so on, becoming an integral function of the multi-media personal computer.

This analysis may suggest that convergence is a simple process of technological substitution. However, analysis of convergence as substitution must also take into account two "emergent phenomena":-

- Computer technology is being changed profoundly by the new interfaces required for telecommunications applications. The transition from keyboard/printer interaction to the new forms of voice and video input and output appropriate to telecommunications applications changes the nature of computing technology.
- The integration of so many human communications media into a single coherent system in itself creates a new technology with opportunities for the development of new products and services.

The next section presents a model of how such new phenomena arise through the learning infrastructure of information technology.

4 The Learning Infrastructure of Information Technology

The number of transistors on a chip has seen a 1,000,000,000 increase in less than 40 years, whereas other high-technology industries have typically seen less than 100 performance increase in 100 years (Gaines, 1991). This improvement depends on the capacity of silicon to support minute semiconductor logic circuits, but this capacity could not have been fully exploited over 9 orders of magnitude performance improvement without the use of the computer to support the design and fabrication of such circuits. This is one example of a positive feedback loop within the evolution of computers, that the computer industry has achieved a learning curve that is unique in its sustained exponential growth because each advance in computer technology has been able to support further advances in computer technology. Such positive feedback is known to give rise to emergent phenomena in biology (Ulanowicz, 1991) whereby systems exhibit major new phenomena in their behavior. The history of computing shows the emergence of major new industries concerned with activities that depend upon, and support, the basic circuit development but which are qualitatively different in their conceptual frameworks and applications impacts from that development; for example, programming has led to a software industry, human-computer interaction has led to an interactive applications industry, document representation has led to a desktop publishing industry, and so on.

Each of these emergent areas of computing has had its own learning curve (Linstone and Sahal, 1976), and the growth of information systems technology overall may be seen as the cumulative impact of a tiered succession of learning curves, each triggered by advances at lower levels and each supporting further advances at lower levels and the eventual triggering of new advances at higher levels (Gaines, 1991). It has also been noted in many disciplines that the qualitative phenomena during the growth of the learning curve vary from stage to stage (Crane, 1972; De Mey, 1982; Gaines and Shaw, 1986). The era before the learning curve takes off, when too little is known for planned progress, is that of the inventor having very little chance of success but continuing a search based on intuition and faith. Sooner or later some inventor makes a *breakthrough* and very rapidly his or her work is *replicated* at research institutions world-wide. The experience gained in this way leads to *empirical* design rules with very little foundation except previous successes and failures. However, as enough empirical experience is gained it becomes possible to inductively model the basis of success and failure and develop *theories*. This transition from empiricism to theory corresponds to the maximum slope of the logistic learning curve. The theoretical models make it possible to *automate* the scientific data gathering and analysis and associated manufacturing processes. Once automaton has been put in place effort can focus on cost reduction and quality improvements in what has become a *mature* technology.

Figure 3 shows a tiered succession of learning curves for information technologies in which a breakthrough in one technology is triggered by a supporting technology as it moves from its research to its empirical stage. Also shown are trajectories shown the eras of *invention*, *research*, *product innovation*, *long-life product lines*, *low-cost products*, and *throw-away products*. One phenomenon not shown on this diagram is that the new industries can sometimes themselves be supportive of further development in the industries on which they depend. Thus, in the later stages of the development of an industrial sector there will be a tiered structure of interdependent industries at different stages along their learning curves.

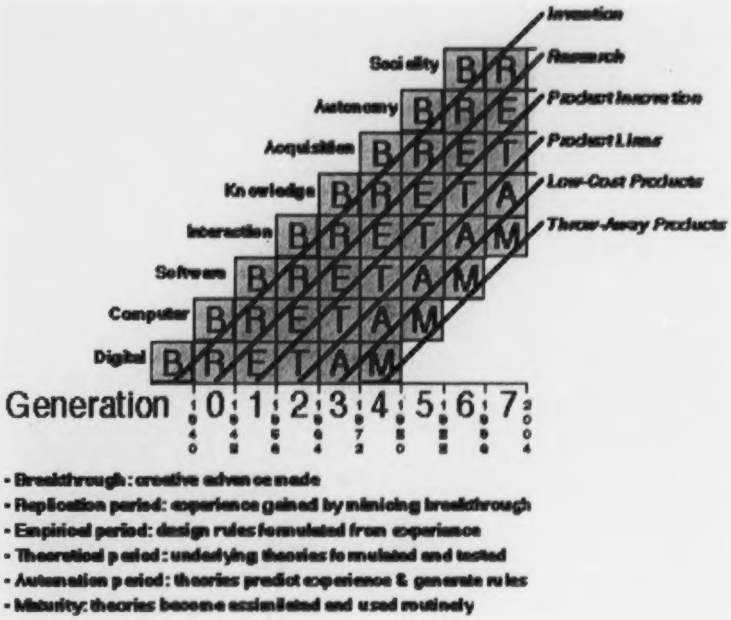

Invention
Research
Product Innovation
Product Lines
Low-Cost Products
Throw-Away Products
Sociality
Autonomy
Acquisition
Knowledge
Interaction
Software
Computer
Digital
B R E T A M
Generation
0 1 2 3 4 5 6 7
1940
1948
1956
1964
1972
1980
1988
1996
2004
• Breakthrough: creative advance made
• Replication period: experience gained by mimicing breakthrough
• Empirical period: design rules formulated from experience
• Theoretical period: underlying theories formulated and tested
• Automation period: theories predict experience & generate rules
• Maturity: theories become assimilated and used routinely

Figure 3 The infrastructure of information technology

The BRETAM tiered learning curves infrastructure of Figure 3 brings together the various phenomena of convergence in an integrated model which has the potential both to explain the past and forecast the future. Well's vision of a world brain immediately predates the initial breakthrough that triggered the learning curves of information technology. The invention of the digital computer was triggered by the type of problem in the processes of civilization that he hoped to prevent. He foresaw a technological solution but not the technologies that actually provided it. Bush, as inventor of the differential analyzer and with his war time knowledge of computing as Roosevelt's advisor, described Memex in the context of the relevant technology. However, it was not until the 1960s that the mean time between failures of computers became long enough to make interactive use routinely possible, and it was not until the 1970s that the costs became low enough for "personal computers" to be developed and these did not come into widespread use until the 1980s.

The relevant learning curves in Figure 3 are the lower four: digital electronics; computer architecture; software; and interaction. The *product innovation* trajectory passes through the last of these in the fourth generation, 1972-1980, and led to the premature development of Viewdata and Videotex products and Martin's detailed forecasts of the potential for a wired society. However, the mass market potential for wired society technology at costs comparable to other mass media such as the telephone and television is dependent on the cost reductions possible in the post-maturity phase of the learning curves leading to *throw-away products*. This trajectory passes through the interaction learning curve in the current seventh generation era, 1996-2004, and it is this that has made the information highway economically feasible.

In projecting the BRETAM model into the future, one critical analysis is whether the learning curves in the lower level technologies can be sustained. The learning curve for the number of devices on a chip has been maintained as a continuing exponential growth even though our knowledge of the underlying silicon device technology is mature because it has been possible to continue to miniaturize the individual transistors using increasingly refined and automated production processes. Such miniaturization is now coming up against fundamental physical limits and against commercial considerations that there are declining economies of scale in fabricating devices with greater numbers of transistors (Ross, 1995). A Forbes survey of eleven leading industry figures in 1996 gave predictions that Moore's law would fail by the year 2005 (Forbes, 1996). However, current digital circuit technology seems sufficient to support the

BEST COPY AVAILABLE

information highway through one or two generations, particularly given the simplicity of the computer architectures currently in use.

The analysis of product opportunities arising from the existence of the information highway involves the upper learning curves of the BRETAM model--knowledge representation and acquisition, autonomy and sociality. Knowledge representation and processing encompasses all the media that can be passed across the web, not just the symbolic logic considered in artificial intelligence studies but also typographic text, pictures, sounds, movies, and the massive diversity of representations of specific material to be communicated. The significance of discourse in the human communities collaborating through the Internet has been underestimated in the stress on `artificial' intelligence in computer research. Knowledge need not be machine-interpretable to be useful, and it can often be machine-processed, indexed and enhanced without a depth of interpretation one might associate with artificial intelligence. The World Wide Web is already a "pool of human knowledge" (Berners-Lee et al., 1994) and the extension of that pool to encompass more and more knowledge is the most significant way of adding value to the web. The problems with this are socio-economic in that much represented knowledge is owned by copyright holders who seek some financial reward before they will offer it to others. Technologically it is important to develop ways of charging for access to knowledge at a low enough rate to encourage widespread use at a high enough volume to compensate the knowledge provider. The knowledge-level problem for the information highway is not so much representation and processing but rather effective trading.

The growth of available material on the web is already causing problems of information overload. The 25 to 50M documents currently available are not readily searched by an individual and the web was not designed for central indexing. The acquisition learning curve is at a level where it has been possible to solve the problem using software programs, spiders, that crawl the web gathering information and indexing it by content so that documents can be retrieved through a key word search. In 1995 indexing spiders solved the problem of acquiring a dynamic model of the rapidly expanding web. However, the simple key word searches offered for retrieval have already become inadequate in that they usually result in a large corpus of documents most of which are irrelevant to the searcher. Information retrieval techniques have to be improved, probably to allow queries to be expressed in natural language and refined through a natural language dialog. Ultimately a user should not be able to determine whether a query is being answered by another person or by a computer program.

5 Conclusions

This article has analyzed the technological infrastructure of convergence to an information highway, tracing the origins of the concept, the false starts, the growth and origins of the Internet and World Wide Web, convergence as a substitution process, and the learning curves of the technologies involved. A number of substitution processes underlying convergence have been identified: electronic for mechanical devices; digital for analog devices; and general-purpose programmable devices for special-purpose devices. A model of convergence in terms of a tiered infrastructure of learning curves in information technology has been proposed and used to explain the past and forecast the future.

Information technology is characterized by high rates of growth in performance parameters sustained over long periods. The number of devices on a chip has grown by 9 orders of magnitude in 37 years. Clock speeds of computers have grown by some 6 orders of magnitude over the same period. The number of computers connected through the Internet is growing at 100% a year and has grown by 7 orders of magnitude in 27 years. The volume of traffic on the Internet is growing at over 100% a year, and that component attributable to the World Wide Web was growing at 1,000% a year when last accurately measurable in 1994.

It is suggested that these high sustained growth rates have been possible because computer technologies are mutually supportive providing positive feedback such that advances in existing technologies trigger breakthroughs in new technologies which themselves help to sustain the advance of the existing technologies. The main problem in forecasting the future of convergence and information technology in general is that the learning curves of most of the major performance parameters still appear to be in their initial exponential growth phase. This makes it impossible to predict the later parts of the curves from past data. For some parameters there are basic physical limitations to existing technologies that indicate that current growth rates cannot be sustained beyond some 10 years. However, there are possibilities for new materials and new architectures that could maintain effective growth rates for the foreseeable future.

Tracking the individual learning curves of the major technologies that comprise the infrastructure of information technology provides a more detailed account of the present and future state-of-the art of the technologies underlying convergence. The base technologies of digital electronics, general-purpose computer architectures, software and interaction are mature and provide solid foundations for *computer science*. The upper technologies of knowledge representation and acquisition, autonomy and sociality, support product innovation and provide the beginnings of foundations for *knowledge science*. Well's dream of a world brain making available all of human knowledge is well on its way to realization and it is in the representation, acquisition, and access and effective application of that knowledge that the commercial potential and socio-economic impact of convergence lies.

Acknowledgments

Financial assistance for this work has been made available by the Natural Sciences and Engineering Research Council of Canada.

References

Berners-Lee, T. (1993). World-Wide Web Talk at Online Publishing 1993. CERN, Geneva. http://info.cern.ch/hypertext/WWW/Talks/OnlinePublishing93/Overview.html.

Berners-Lee, T. and Cailliau, R. (1990). WorldWideWeb: Proposal for a Hypertext Project. CERN, Geneva. http://info.cern.ch/hypertext/WWW/Proposal.html.

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H.F. and Secret, A. (1994). The World-Wide Web. **Communications ACM 37**(8) 76-82.

Bianculli, D. (1992). **Teleliteracy: Taking Television Seriously**. New York, Simon & Schuster.

Braun, E. and Macdonald, S. (1978). **Revolution in Miniature**. Cambridge, UK, Cambridge University Press.

Bush, V. (1945). As we may think. **Atlantic Monthly 176** 101-108.

CERN (1994). History to date. CERN, Geneva. http://info.cern.ch/hypertext/WWW/History.html.

Chorafas, D.N. (1981). **Interactive Videotex: the Domesticated Computer**. New York, Petrocelli.

Crane, D. (1972). **Invisible Colleges: Diffusion of Knowledge in Scientific Communities**. Chicago, University of Chicago Press.

De Mey, M. (1982). **The Cognitive Paradigm**. Dordrecht, Holland, Reidel.

Fedida, S. and Malik, R. (1979). **The Viewdata Revolution**. London, UK, Associated Business Press.

Forbes (1996). Chips triumphant. **Forbes ASAP February 26** 53-82.

Gaines, B.R. (1971). Through a teleprinter darkly. **Behavioural Technology 1**(2) 15-16.

Gaines, B.R. (1991). Modeling and forecasting the information sciences. **Information Sciences 57-58** 3-22.

Gaines, B.R. and Shaw, M.L.G. (1986). A learning model for forecasting the future of information technology. **Future Computing Systems 1**(1) 31-69.

Goodman, H.J.A. (1987). The "world brain/world encyclopaedia" concept: its historical roots and the contributions of H.J.A. Goodman to the ongoing evolution and implementation of the concept. **ASIS'87: Proceedings 50th Annual Meeting American Society Information Science**. pp.91-98. Medford, New Jersey, Learned Information.

Gruenberger, F., Ed. (1967). **The Transition to On-Line Computing**. Washington, DC, Thompson.

Linstone, H.A. and Sahal, D., Ed. (1976). **Technological Substitution: Forecasting Techniques and Applications**. New York, Elsevier.

Martin, J. (1978). **The Wired Society: A Challenge for Tomorrow**. Englewood Cliffs, New Jersey, Prentice-Hall.

Mauchly, J.W. (1942). The use of high speed vacuum tube devices for calculating. Randell, B., Ed. **The Origins of Digital Computers: Selected Papers**. pp.329-332. Springer, Berlin.

NW (1996). Internet Domain Survey. Network Wizards. http://www.nw.com.

Orr, W.D., Ed. (1968). **Conversational Computers**. New York, Wiley.

Parkhill, D.F. (1966). **The Challenge of the Computer Utility**. Reading, MA, Addison-Wesley.

Ross, P.E. (1995). Moore's second law. **Forbes March 25** 116-117.

Rubinoff, M., Ed. (1965). **Toward a National Information System**. Washington D.C., Spartan Books.

Salus, P. (1995). **Casting the Net: From ARPANET to INTERNET and Beyond**. Reading. MA, Addison-Wesley.

Sigel, E. (1980). **Videotext: The Coming Revolution in Home/Office Information Retrieval**. White Plains, NY, Knowledge Industry Publications.

Ulanowicz, R.E. (1991). Formal agency in ecosystem development. Higashi, M. and Burns, T.P., Ed. **Theoretical Studies of Ecosystems: The Network Perspective**. pp.58-70. Cambridge, Cambridge University Press.

Wells, H.G. (1938). **World Brain**. New York, Doubleday.

Winsbury, R. (1981). **Viewdata in Action: a Comparative Study of Prestel**. New York, McGraw-Hill.

Woolfe, R. (1980). **Videotex: the New Television/Telephone Information Services**. London, Heyden.

KSI Page

gaines@cpsc.ucalgary.ca 16-Jul-96

Utility, Usability and Likeability: Dimensions of the Net and Web

Brian R. Gaines, Mildred L. G. Shaw and Lee Li-Jen Chen

Knowledge Science Institute
University of Calgary
Alberta, Canada T2N 1N4
{gaines, mildred, lchen}@cpsc.ucalgary.ca
http://ksi.cpsc.ucalgary.ca/KSI/

Abstract: The explosive growth of the Internet and World Wide Web are well substantiated. The growth has no central origin but is a distributed social phenomenon emerging from the independent perceptions of many individuals and communities that the net and web have a high utility in attaining their objectives. However, the very loosely coordinated growth of net technologies have led to significant human factors problems, and modeling diffuse communities interacting through the net requires new human factors methodologies and theories. This article analyses the human factors of the net and web in terms of the standard dimensions of utility, usability and likeability.

- 1 Introduction
- 2 A Layered Protocol Model of the Human Factors of the Internet
- 3 Defining Communities on the Internet and World Wide Web
- 4 Dimensions of the Utility of the Internet and World Wide Web
- 5 An Example of a Usability Bug--Replying to a List
- 6 Conclusions
- References

1 Introduction

Shackel (1991) factors human factors issues along three major dimensions:-

- *utility*--will it do what is needed functionally?
- *usability*--will the users actually work it successfully?
- *likeability*--will the users feel it is suitable?

Shackel measures usability on four dimensions:

- *effectiveness*--performance in accomplishment of tasks--the access to potential utility
- *learnability*--degree of learning to accomplish tasks--the effort required to access utility
- *flexibility*--adaptation to variation in tasks--the range of tasks for which there is utility
- *attitude*--user satisfaction with system--the manifestation of potential likeability

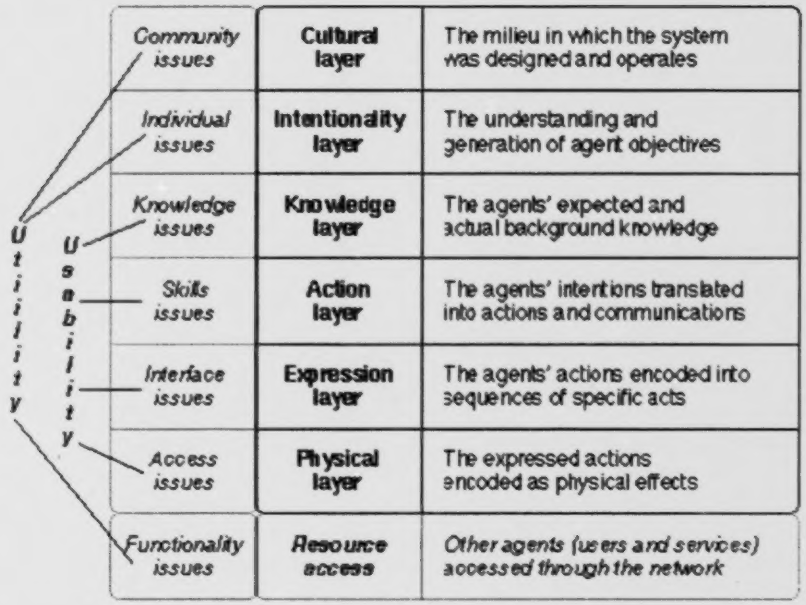
In applying this analysis to the net and web one has to take into account that *utility* in a collaborative environment is not just for individual users but for communities. For example, the *accessibility* of a system to all members of the relevant reference community is a utility consideration. One has to take into account that *usability* is not intensionally defined in terms of compliance with human factors guidelines, but rather extensionally defined in terms of evidence of a high proportion of effective users. *Likeability* is a critical factor to user adoption of a technology, particularly in a competitive market place, but it is sometimes taken as a "subjective" dimension not subject to formal modeling. Trevino and Webster (1992) developed a formal model of likeability in the context of computer-mediated communication (CMC) using Csikszentmihalyi's (1990) concept of *flow* underlying the psychology of optimal experience, and Hoffman and Novak (1995) have recently applied it to the analysis of marketing through the web.

It is important to analyze the utility of a technology before one considers its usability. The analysis of what is "needed functionally" is fundamental to understanding the problems users have in fulfilling that need. The utility of email, for example, often appears so obvious that human factors analysis focuses only on the user interface. However, a deeper analysis of needed functionality will reveal that the support of discourse through a textual medium with no voice intonation, body language, and so on, involves complex human factors (Walther, 1992; Spears and Lea, 1994). The fact that the computer interface itself does not specifically support the `netiquette' (Shea, 1994) aspects of electronic discourse does not mean that it is a human factors problem that can be neglected in system design. Minimally, it draws attention to the need for user guidance or training, and technically it suggests that tools that evaluate email in terms of the emotional loading of the words involved might be useful to users. Human factors studies of the net and web have to analyze the social expectations of the community within which the communication takes place, and the users' objectives in communicating. The computer system has no technical `understanding' of these issues, but they determine the utility of a computer-mediated system as much as the lower layers, and failure by users to conform with the cultural norms or express their intentions effectively is failure of usability as much as is pressing the wrong icon to send the mail.

2 A Layered Protocol Model of the Human Factors of the Internet

The human factors issues discussed above can be analyzed in terms of the layered protocol models which have been developed to provide a conceptual framework for complex systems of people and computing services (Gaines, 1988). Figure 1 shows a 6 layer model that partitions each sub-system, person or computing service, into:

- A *cultural* layer which encompasses the milieu of the community within which the system operates and influences user intentions;
- An *intentionality* layer which encompasses the objectives of the agents operating within the system;
- A *knowledge* layer which encompasses the overt background knowledge necessary to translate agent intentions into actions;
- An *action* layer which encompasses the skills or services of the agents as they translate intentions into actions and communications;
- An *expression* layer which encompasses the encoding of actions or communications into a specific sequence of discrete acts;
- A *physical* layer which encompasses the encoding of actions into physical effects.



Community issues	**Cultural layer**	The milieu in which the system was designed and operates
Individual issues	**Intentionality layer**	The understanding and generation of agent objectives
Knowledge issues	**Knowledge layer**	The agents' expected and actual background knowledge
Skills issues	**Action layer**	The agents' intentions translated into actions and communications
Interface issues	**Expression layer**	The agents' actions encoded into sequences of specific acts
Access issues	**Physical layer**	The expressed actions encoded as physical effects
Functionality issues	***Resource access***	*Other agents (users and services) accessed through the network*

Figure 1 Layered protocol model for human-computer interaction

The action and expression layers are separated to allow actions to be conceptualized both as unitary abstractions and as the particular expression of those abstractions as a sequence of acts. For example auto-dialing may be conceived as a single psychological entity even though it may be expressed by a menu selection, dialog box click or command line text entry. The term `agent' is used to cover both human users and computer services. At the current state of the art, the top three layers are design time considerations in developing computer services and only the three low levels are operational at run time, although the development of `knowledgeable agents' as a computer technology is a major research area.

Multiple 6 layer models of users and 3 layer models of services may be combined to provide a model of a complex system in which users communicate with other users and services through a network at the physical layer, establishing `virtual circuits' at the upper layers (Gaines, 1988). This is shown as "resource access" at the bottom of Figure 1 which subsumes the functionality that the agent can access through the network.

Shackel's *utility* can now be seen as an evaluation of the extent to which a user's intentions can be realized through processes accessing functionality available. For the net and web it is also appropriate to assess functionality in relation to a community's needs.

Shackel's *usability* can be seen as an evaluation of the extent to which users can translate their intentions into effective actions to access the functionality. It factors through 4 layers into:-

- Knowledge issues concerned with the background knowledge that the user has available through experience or training;
- Skills issues concerned with the user's capability to translate intentions into actions using the background knowledge as appropriate;
- Interface issues concerned with the facilities provide for the user to translate abstract actions into a sequence of acts;
- Access issues concerned with the provisions for the user to access the functionality specified through the sequence of acts.

Shackel's *learnability, flexibility and likeability* are non-functional aspects of usability that are manifest in each layer. Likeability is the least studied of the three because it has seemed a subjective matter with no behavioral model that can only be tested through questionnaire techniques. However, in recent years

Csikszentmihalyi's (1990) concept of *flow* as the phenomenon underlying the psychology of optimal experience has been applied to modeling user satisfaction with CMC (Trevino and Webster, 1992). In their application of the model to flow phenomena in interaction with the web, Hoffman and Novak (1995) summarize the concept as:-

> *"Flow has been described as `the process of optimal experience' achieved when a sufficiently motivated user perceives a balance between his or her skills and the challenges of the interaction, together with focused attention. Flow activities in the Web, specifically network navigation, facilitate concentration and involvement because they are distinct from the so-called `paramount reality' of everyday existence."*

It is reasonable to propose that likeability correlates with a flow state in which a motivated user undertakes a task whose level of difficulty is at some particular level that suits their individual needs. Too low a level results in boredom and too high a level in anxiety, and the optimal level results in the intense satisfaction with the activity that Csikszentmihalyi terms flow.

In the flow model, likeability is not associated with a particular layer of the protocol but rather with the appropriate level of activity involving the protocol. This explains some of the paradoxical aspects of usability analysis--high usability does not imply high likeability, and many well-liked interfaces and systems are poor from a usability standpoint. From a flow perspective, a simple task which is boring through a highly usable interface may be enhanced in its likeability by decreasing the interface usability to present a greater challenge to the user.

Similar considerations apply to flexibility which generally increases utility but may decrease usability because it involves the use of the system in modes for which it was not designed. The dimensions of human factors evaluation are not monotonically related and there are generally trade-offs between them.

3 Defining Communities on the Internet and World Wide Web

The cultural layer at the top of Figure 1 itself requires partitioning when one considers the different forms of community supported through CMC. It is already been noted that at least three types of community need to be distinguished: the highly-coordinated, goal-directed *teams*; the more loosely coordinated *special-interest communities*, such as professional sub-disciplines; and the largely uncoordinated *Internet world at large* whose members have in common only the use of CMC. Figure 2 shows the layered protocol model extended to groups of agents grouped as specified.

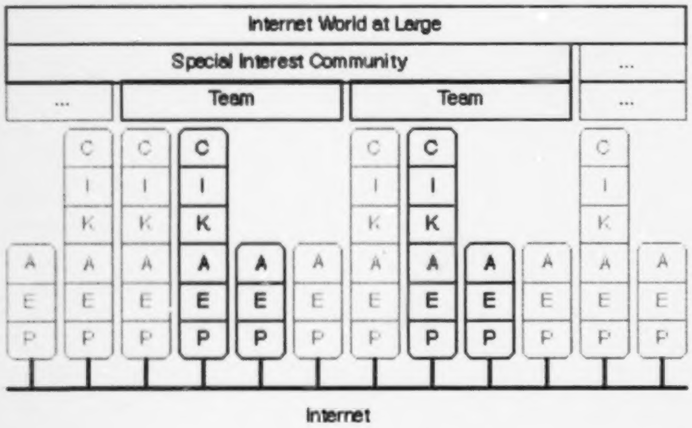


Figure 2 Communities within the layered protocol model
(3 layers represent computer services and 6 layers users, gray items represent an indefinite iteration)

For purposes of human factors analysis one needs a precise behavioral definition of these various form of community. How may they be differentiated conceptually and through empirical observation? One can regard a community as a set of individuals that provide resources to one another with the most significant dimension relating to the coordination of the community being that of the *awareness* of who is providing a particular resource and who is using it. In the tightly-coupled team, each person is usually aware of who will provide a particular resource and often of when they will provide it. In logical terms, this can be termed *extensional awareness* because the specific resource and provider are known, as contrasted to *intensional awareness* in which only the characteristics of suitable resources or providers are known.

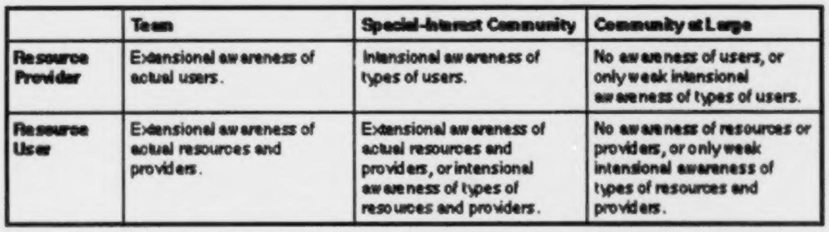
A team can be treated from a *collective stance* (Gaines, 1994) as a single psychological individual that behaves as a compound role generated by the distributed activities of roles in a number of people. Each resource provider in a team has an *extensional awareness* of their actual resource users, and each resource user has an *extensional awareness* of the resource and who will provide it.

In a special interest community resource providers usually do not have such extensional awareness of the resource users, and, if they do, can be regarded as forming teams operating within the community. Instead, resource providers usually have an *intensional awareness* of the resource users in terms of their characteristics as *types* of user within the community. The classification of users into types usually corresponds to social norms within the community, such as the ethical responsibilities in a professional community to communicate certain forms of information to appropriate members of the community. Resource users in a special interest community may have an extensional awareness of particular resources or resource providers, or an intensional awareness of the types of resource provider likely to provide the resources they require. This asymmetry between providers and users characterizes a special interest community and also leads to differentiation of the community in terms of core members of whom many users are extensionally aware, and sub-communities specializing in particular forms of resource.

In the community of users at large, there is little awareness of particular resources or providers and only a general awareness of the rich set of resources is available. Awareness of the characteristics of resources and providers is vague, corresponding to *weak intensional awareness*.

These distinctions lead to human factors consequences in terms of the appropriate awareness mechanisms that need to be established on the network for the communities to function. In a team, resources may be identified precisely by location and name. In a special-interest community, resources may be identified by an intensional indexing scheme that classifies them in terms of the distinctions made by that community. In the community of users at large, resources may be identified both by indexing their content type using a wide variety of taxonomies and by indexing their actual content.

These distinctions are summarized in Figure 3, and it is clear that the classification of awareness can lead to a richer taxonomy of communities than the 3-way division defined. Analysis of awareness in these terms allows the structure of a community to be specified in operational terms, and in complex communities there will be complex structures of awareness. The coarse divisions into sub-teams and sub-special interest communities provides a way of reducing this complexity in modeling the community.



	Team	**Special-Interest Community**	**Community at Large**
Resource Provider	Extensional awareness of actual users.	Intensional awareness of types of users.	No awareness of users, or only weak intensional awareness of types of users.
Resource User	Extensional awareness of actual resources and providers.	Extensional awareness of actual resources and providers, or intensional awareness of types of resources and providers.	No awareness of resources or providers, or only weak intensional awareness of types of resources and providers.

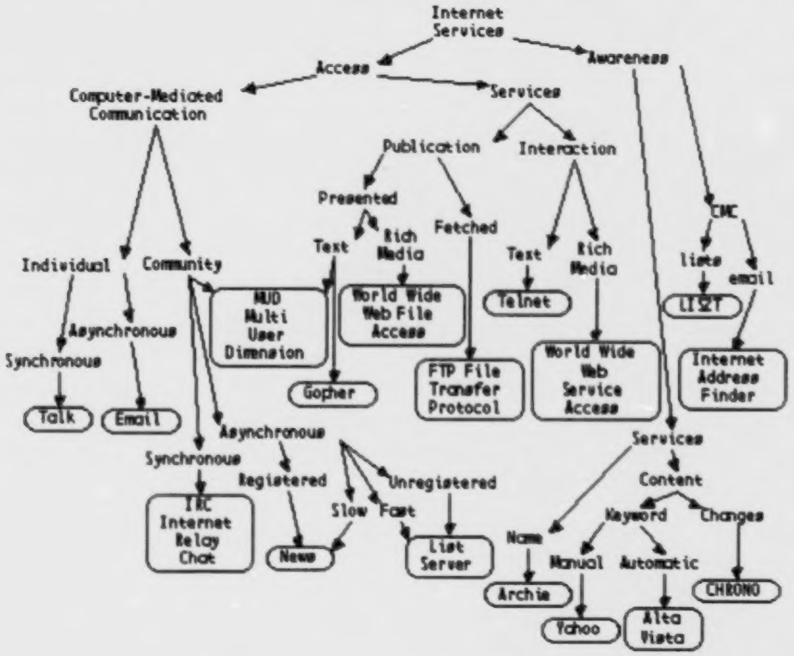
Figure 3 Communities distinguished by awareness

BEST COPY AVAILABLE

The differentiation of communities in terms of awareness draws attention to the significance of supporting various aspects of awareness in a CMC system. As noted above *resource awareness*, the awareness that specific resources or resources with specified characteristics exists, may be supported by various indexing and search procedures. However, there is also a need to support *chronological awareness* (Chen and Gaines, 1996), the awareness that a resource has changed or come into existence. These are major human factors issues in users coping with the net as a rich and rapidly changing system of resources.

4 Dimensions of the Utility of the Internet and World Wide Web

In examining the utility of the net and web it is useful to classify all the major services in terms of the significant distinctions that determine their relative utilities as shown in Figure 4. It characterizes the major net services in terms of their utility for CMC, resource access or search. It sub-classifies CMC in terms of whether it is individual-to-individual discourse or community discourse; synchronous with the participants conversing in real time or asynchronous with substantial time delays in responses. It sub-classifies asynchronous community discourse by whether the channel is slow or fast, and whether the community is centrally registered or not. It sub-classifies resource access in terms of whether it is: a static resource or an interactive service; just fetched or presented when fetched; text or rich media. It sub-classifies search by whether it is: by resource name or content; by keywords or by change in contents; and whether the keywords are generated manually or automatically.



Internet Services
Access
Awareness
Computer-Mediated Communication
Services
Publication
Interaction
Presented
Fetched
Text
Rich Media
Text
Rich Media
CMC
lists
email
Individual
Community
MUD Multi User Dimension
World Wide Web File Access
Telnet
LISZT
Asynchronous
Synchronous
Internet Address Finder
FTP File Transfer Protocol
World Wide Web Service Access
Gopher
Talk
Email
Asynchronous
Synchronous
Services
Registered
Unregistered
Content
IRC Internet Relay Chat
Slow
Fast
Keyword
Changes
Name
News
List Server
Manual
Automatic
Archie
CHRONO
Yahoo
Alta Vista

Figure 4 Distinctions characterizing the utility of various Internet services

The major services classified are:-

- *Talk*, the facility for one user to send a message directly to the terminal of another user. This provides individual, synchronous computer-mediated communication.
- *Email*, the facility for one user to send a message to the mailbox of another user. This provides individual, asynchronous computer-mediated communication.
- *Internet Relay Chat (IRC)*, the facility for a user to join a chat group and send a message directly to the terminals of the group. This provides community, synchronous computer-mediated communication.

- *News*, the facility for a user to mail a message to a registered newsgroup archive and to access messages in the archive. This provides community, asynchronous computer-mediated communication. Because the archives are maintained on a local server and updated through a chain of servers the updating is slow, possibly taking several days.
- *List Server*, the facility for a user subscribe to a list server and mail a message to it which it mails to all members on the list. This again provides community, asynchronous computer-mediated communication. Because the mailing to the list is fast (except for moderated groups where the mail is manually checked), list servers provide more interactive discourse than news groups. However, the registration of news groups makes them easier to discover, and, for high-volume discourse, users may prefer that it is not posted to their mailbox.
- *Multi-User Dimension (MUD)*, the facility for a user to `enter a dimension', communicate directly with others there, and leave and retrieve documents. This provides community, computer-mediated communication and text resource access.
- *Gopher*, the facility for a user to retrieve a text document from a hierarchically structured archive.
- *World Wide Web file access*, the facility for a user to retrieve multi-media documents from an archive through hypertext links embedded in the document.
- *Telnet*, the facility for a user to interact with a remote machine through a console window providing a command line interface. This provides remote interactive access to services based on text.
- *World Wide Web service access*, the facility for a user to enter information into an HTML form and transmit it to a remote server. This provides remote interactive access to services based on rich media.
- *File Transfer Process (FTP)*, the facility for a user to retrieve a file by site and name. This provides general file access but FTP clients generally lack the capability to present the files retrieved.
- *Internet Address Finder*, a service for a user to search for the email address of a person by their name. One of the problems of the net is the lack of an overall directory of users.
- *LISZT*, a service for a user to search for a list server by its name. This attempts to overcome the problem that there is no central directory of list servers.
- *Archie*, a facility for a user to search the net for files by name. This provides search facilities for files with known names.
- *Yahoo*, a facility for a user to search the net for resources by name and key word through a manually entered classification. This provides search facilities for resources specified by their name or type.
- *Alta Vista*, a facility for a user to search the net for resources by content. This provides search facilities for resources with specified content.
- *CHRONO*, a facility for a user to search a site through a list of changed resources in reverse chronological order. This provides search facilities for resources by recency.

5 An Example of a Usability Bug--Replying to a List

Significant usability problems have arisen in existing Internet technology. For example, a usability problem has been introduced into the operation of list servers by the use of email browsers to access them. Because of the length and complexity of the information, mail browsers generally show only abbreviated headers which are adequate for normal mail but problematic for mail from list servers. This is because the servers use a simple trick to allow a reply to a message from the list to be addressed to the list rather than to the originator. The Internet mail protocol allows both a "From:" field and a "Reply-To:" field to be specified in the mail header. In normal mail usage the "Reply-To:" field is absent and a reply is sent to the address in the "From:" field. In list servers the "From:" field is filled with the originator and the "Reply-To:" field is filled with the list server address so that replies go to the list. However, email browsers generally show the "From:" field as the origin of the mail, and it appears to the user as if they are replying to the individual who sent the mail to the list. This often leads to embarrassingly personal messages intended only for one person being mailed to the entire list. It would be better if email browsers were configured to show the "Reply-To:" field in preference to the "From:" field, perhaps indicating that the "From:" field is different.

Some list administrators set the "Reply-To" field to be the originator to avoid this problem, but this reduces the usability of the interface since most replies are intended for the community, and having to enter the list server address in a reply is an impediment to spontaneous discourse. Some list administrators overcome the problem by "moderating" their lists and examining each item of mail before authorizing it to go to the list. This is again a serious impediment to discourse since it introduces delays, taking the list turnaround time from minutes to days.

The usability problems created by email browsers' failure to represent the "Reply-To:" field adequately, and by attempts to fix this, may be seen as a conflict between the *knowledge* and *skills* layers in the layered protocol model. The user is required to disrupt the skilled activity of discourse by a knowledge-driven override. Email discourse is similar to vocal discourse in that when a person receives a message they may instantly conceive a reply. The emission of the reply vocally is mediated automatically without disruption of the chain of consciousness framing the reply. Hitting the "Reply" key in an email browser is a similar subconscious reaction made automatically without disruption of the process of composing a reply. However, having to remember or ascertain whether the reply is going to the originator or the list, and manually fix the address if it is not what is wanted, disrupts the composition of a reply.

The problems may also be seen as arising from the *flexibility* of the email browsers which enables them to be used as list server browsers. Flexibility is generally positive for utility because it widens the range of applications, but it is often negative for usability in that it involves operating the system outside the range of situations for which it was designed to be usable.

6 Conclusions

This article presents a framework for representing and analysing the human factors of the Internet and World Wide Web that applies the classical analysis in terms of utility, usability and likeability to the social environment of the net and web. A layered protocol model introduces a spectrum of distinctions ranging from culture through knowledge and action to physical access. Various forms of community are defined operationally in terms of types of awareness between resource providers and users. The dimensions of utility are used to provide a taxonomy of Internet services. Paradoxical trade-offs between utility, usability and likeability are illustrated as are some usability problems of existing Internet services.

Acknowledgments

This work was funded in part by the Natural Sciences and Engineering Research Council of Canada.

References

Chen, L.L.-J. and Gaines, B.R. (1996). Methodological issues in studying and supporting awareness on the World Wide Web. Maurer, H., Ed. **Proceedings of WebNet96**. pp.this conference. Charlottesville, VA, Association for the Advancement of Computing in Education.

Csikszentmihalyi, M. (1990). **Flow: The Psychology of Optimal Experience**. New York, Harper and Row.

Gaines, B.R. (1988). A conceptual framework for person-computer interaction in distributed systems. **IEEE Transactions Systems, Man & Cybernetics SMC-18**(4) 532-541.

Gaines, B.R. (1994). The collective stance in modeling expertise in individuals and organizations. **International Journal of Expert Systems** 7(1) 21-51.

Hoffman, D.L. and Novak, T.P. (1995). Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations. Owen Graduate School of Management, Vanderbilt University, Nashville, TN.

http://www2000.ogsm.vanderbilt.edu/cmepaper.revision.july11.1995/cmepaper.html.

Shackel, B. (1991). Usability--context, framework, definition, design and evaluation. Shackel, B. and Richardson, S., Ed. **Human Factors for Informatics Usability**. pp.21-37. Cambridge, UK, Cambridge University Press.

Shea, V. (1994). **Netiquette**. San Francisco, Albion Books.

Spears, R. and Lea, M. (1994). Panacea or panopticon? The hidden power in computer-mediated communication. **Communication Research 21**(4) 427-459.

Trevino, L.K. and Webster, J. (1992). Flow in computer-mediated communication. **Communication Research 19**(5) 539-575.

Walther, J.B. (1992). Interpersonal effects in computer-mediated communication: A relational perspective. **Communication Research 19**(1) 52-90.

KSI Page

gaines@cpsc.ucalgary.ca 16-Jul-96

Information Fusion with ProFusion*

Susan Gauch, Guijun Wang
Department of Electrical Engineering and Computer Science
The University of Kansas, Lawrence, KS 66045, USA
{sgauch, gwang}@eecs.ukans.edu
*http://www.designlab.ukans.edu/ProFusion.html

Abstract: The explosive growth of the World Wide Web, and the resulting information overload, has led to a mini-explosion in World Wide Web search engines. This mini-explosion, in turn, led to the development of ProFusion, a meta search engine. Educators, like other users, do not have the time to evaluate multiple search engines to knowledgeably select the best for their uses. Nor do they have the time to submit each query to multiple search engines and wade through the resulting flood of good information, duplicated information, irrelevant information, and missing documents. ProFusion sends user queries to multiple underlying search engines in parallel, retrieves and merges the resulting URLs. It identifies and removes duplicates and creates one relevance-ranked list. If desired, the actual documents can be pre-fetched to remove yet more duplicates and broken links. ProFusion's performance has been compared to the individual search engines and other meta searchers, demonstrating its ability to retrieve more relevant information and present fewer duplicates pages. Future developments include analyzing the documents for improved ranking, automatically submitting queries to the most appropriate search engines, and modifying ProFusion to be an information filtering and dissemination system.

1. Introduction

There are a huge number of documents on the World Wide Web, making it very difficult to locate information that is relevant to a user's interest. Search tools such as InfoSeek[InfoSeek] and Lycos[Lycos] index huge collections of Web documents, allowing users to search the World Wide Web via keyword-based queries. Given a query, such search tools search their individual index and present the user with a list of items that are potentially relevant, generally presented in ranked order. However large the indexes are, still each search tool indexes only a subset of all documents available on WWW. As more and more search tools become available, each covering a different (overlapping) subset of Web documents, it becomes increasingly difficult to choose the right one to use for a specific information need. ProFusion has been developed to help deal with this problem.

2. Related Work

There are several different approaches to managing the proliferation of Web search engines. One solution is to use a large Web page that lists several search engines and allows users to query one search engine at a time. One example of this approach is All-in-One Search Page [Cross]. Unfortunately, users still have to choose one search engine to which to submit their search.

Another approach is to use intelligent agents to bring back documents that are relevant to a user's interest. Such agents [Balabanovic et al. 1995][Knoblock et al. 1994] provide personal assistance to a user. For example, [Balabanovic et al. 1995] describes an adaptive agent that can bring back Web pages of a user's interest daily. The user gives relevance feedback to the agent by evaluating Web pages that were brought back. The agent them makes adjustment for future searches on relevant Web pages. However, these agents [Balabanovic et al. 1995][Knoblock et al. 1994] gather information from only their own search index, which may limit the amount of information they have access to.

A different approach is the meta search method which builds on top of other search engines. Queries are submitted to the meta search engine which in turn sends the query to multiple single search engines. When retrieved items are returned by the underlying search engines, it further processes these items and presents relevant items to the user. ProFusion [ProFusion], developed at the University of Kansas, is one such search engine.

The idea of using a single user interface for multiple distributed information retrieval systems is not new. Initially, this work concentrated on providing access to distributed, heterogeneous database management systems [Arens et al. 1993]. More recently, meta searchers for the WWW have been developed. For example, SavvySearch [Dreilinger] selects the most promising search engines automatically and then sends the user's query to the selected search engines (usually 2 or 3) in parallel. SavvySearch does very little post-processing. For example, the resulting document lists are not merged. MetaCrawler [Selberg et al. 1995][MetaCrawler], on the other hand, sends out user's query to all search engines it handles and collates search results from all search engines. What distinguishes ProFusion from others is that it uses sophisticated yet computationally efficient post-processing.

3. ProFusion

3.1 General Architecture

ProFusion accepts a single query from the user and sends it to multiple search engines in parallel. The current implementation of ProFusion supports the following search engines: Alta Vista [Alta Vista], Excite [Excite], InfoSeek [InfoSeek], Lycos [Lycos], Open Text [Open Text], and WebCrawler [WebCrawler]. By default, ProFusion will send a query to InfoSeek, Lycos, and Excite, but the user may select any or all of the supported search engines. If the user prefers, the system will analyze the user's query, classifying it into a subject or multiple subjects. Based on this analysis, the system will automatically pick the top three search engines that perform best on this subject or these subjects. However the search engines are selected, the search results they return are then further processed by ProFusion. The post-processing includes merging the results to produce a single ranked list, removing duplicates and dead references, and pre-fetching documents for faster viewing and further analysis.

3.2 User Interface

ProFusion queries are simple to form; they are merely a few words describing a concept. Online help is available via a Help button that leads users to a page explaining the query syntax, including sample queries. Users need only enter a query and press the "Search" button, however there are several options available which give the user more control over their search. The first option specifies whether or not the user wants to have a short summary displayed for each retrieved item. The benefit of displaying retrieved items without a summary is that a user can more quickly scan retrieved items by title. The second option allows users to manually select the search engine(s) to which their query is sent, or to have the system choose automatically (described in Section 3.1). If the user is selecting the search engines, they may choose any number of search engines from one to all six. When "Automatic Pick Best 3" is selected, the system to selects the best three search engines based on the words in the query.

3.3 Duplicate Removal

Since the underlying search engines overlap in the Web pages they index, it is highly likely that they will return some of the same pages in response to a given query. ProFusion attempts to remove these duplicated pages, using a few simple rules. The simplest case is when the identical URL has been returned by multiple search engines. Clearly, if two items have exactly the same URL, they are duplicates. More complex rules are necessary to handle the case where the identical page is referenced by slight variations on the same address. For example, the URLs is "http://server/" and is "http://server/index.html" reference the identical page. Handling the previous two cases removes approximately 10 - 20% of the retrieved URLs. However, duplicates may also occur because multiple copies of the same page may exist at different locations. Thus, if two items have different URLs but the same title, they might be duplicates. In this case, we break a URL into three parts: protocol, server, and path. We then use n-gram method to test the similarity of two paths. If they are sufficiently similar, we consider them as duplicates. This appears to work very well in practice, removing an additional 10 - 20% of the URLs, but runs the risk that the URLs point to different versions of the same document, where one is more up-to-date than the other. To avoid this risk, we could retrieve the potential duplicates in whole or in part, and then compare the two documents. However, this would increase network traffic and might be substantially slower. This capability has been developed, and will soon be added as an option.

3.4 Merge Algorithms

How to best merge individual ranked lists is an open question in searching distributed information collections [Voorhees et al. 1994]. Callan [Callan et al. 1995] evaluated merging techniques based on rank order, raw scores, normalized statistics, and weighted scores. He found that the weighted score merge is computationally simple yet as effective as a more expensive normalized statistics merge. Therefore, in ProFusion, we use a weighted score merging algorithm which is based on two factors: the value of the query-document match reported by the search engine and the estimated accuracy of that search engine.

For a search engine *i*, we calculated its confidence factor, CFi, by evaluating its performance on a set of over 25 queries. The CFi reflects the number of total relevant documents in top 10 hits and the ranking accuracy for those relevant documents. Based on the results, the search engines were assigned CFis ranging from 0.75 to 0.85. More work needs to be done to systematically calculate and update the CFis, particularly developing CFis which vary for a given search engine based on the domain of the query.

When a set of documents is returned by search engine *i*, we calculate the match factor for each document *d*, Mdi, by normalizing all scores in the retrieval set to fall between 0 and 1. We do this by dividing all values by the match value reported for the top ranking document. If the match values reported by the search engine fall between 0 and 1, they are unchanged. Then, we calculate the relevance weight for each document *d*, Rdi, by multiplying its match factor, Mdi, by the search engines confidence factor, CFi. The document's final rank is then determined by merging the sorted documents lists based on their relevance weights, Rdi. Duplicates are identified during the merging process. When duplicates are removed, the surviving unique document's weight is set to the maximum Rdi value of all the copies.

3.5 Search Result Presentation

The merge process described in the previous section yields a single sorted list of items, each composed of a URL, a title, a relevance weight, and a short summary. These items are then displayed to the user in sorted order, with or without the summary, depending on user's preference.

3.6 Other Implementation Details

ProFusion is written in Perl and is portable to any Unix platform. It contains one Perl module for each search engine (currently six) which forms syntactically correct queries and parses the search results to extract each item's information. Other modules handle the user interface, the document post-processing, and document fetching. Due to it's modular nature, it is easy to extend ProFusion to additional search engines.

ProFusion's main process creates multiple parallel sub-processes, and each sub-process sends a search request to one search engine and extracts information from the results returned by the search engine. The main process begins post-processing when all sub-processes terminate by returning their results or by timing out (60 seconds in the current prototype).

3.7 Performance Evaluation

We invited every student in our Spring 1996 Information Retrieval class to select a query he/she was interested in. They then were asked to perform a search on that query using each of 9 search engines: the six underlying search engines used by ProFusion (Alta Vista, Excite, InfoSeek, Lycos, Open Text, WebCrawler); ProFusion; and two other meta search engines (MetaCrawler and Savvy Search). Each participant provided relevance judgments for the top 20 retrieved items from each search engine, noting which were broken links and which were duplicates. The performance of each of the search engines was then compared by accumulating the information on the number of relevant documents, the number of irrelevant documents, the number of broken links, the number of duplicates, the number of unique relevance documents, and the precision. Here, precision = number of unique relevant documents divided by total number of documents retrieved (20 documents in this evaluation). The following is a summary of the results from the 12 independent queries, evaluated by the top 20 retrieved documents (total 240 documents evaluated for each search engine).

	Total	Total	Total	Total	Average
Search Engines	number of relevant (_)	number of irrelevant (_)	Broken links ()	number of unique relevant document (_)	Precision number unique / 240
Single Search Engines					
Alta Vista	108	101	31	99	0.41
Excite	129	104	7	122	0.51
InfoSeek	99	125	16	87	0.36
Lycos	119	104	17	93	0.39
Open Text	72	136	32	54	0.23
WebCrawler	92	130	18	73	0.30
Meta Search Engines					
MetaCrawler	98	118	24	85	0.35
Savvy	127	84	29	112	0.47
ProFusion	142	85	13	134	0.56

Table 1: Performance Comparison

From this table, we see that ProFusion achieved the best average precision of all 9 search engines, since it returned the most relevant documents. We attribute this performance to our sophisticated yet efficient merging algorithm, combined with the removal of duplicates. When more of the documents in the top 20 are unique, there is a better chance that more of them are relevant. ProFusion did a better job in duplicate removal than Savvy and MetaCrawler. ProFusion has 142-134=8 duplicates among 142 relevant documents (5.6%), whereas Savvy Search has 127-112=15 duplicates among 127 relevant documents (11.8%) and MetaCrawler has 98-85=13 duplicates among 98 relevant documents (13.3%). Similar numbers were observed for duplicates among irrelevant retrieved documents. The percentage of broken links retrieved by ProFusion was also lower than any system except Excite.

4. Future Work

Enhancements that are underway include analyzing the retrieved documents to improve the ranking, incorporating user preferences (e.g., do they prefer content-bearing pages which contain mostly text or summary pages which primarily contain links to further pages), and improving the automatic search engine selection. In addition, we plan to add the ability to automatically rerun searches on a periodic basis, presenting only new or updated URLs to the user. This will provide a personal search assistant/information filtering capability.

Acknowledgments

This work was funded by the University of Kansas General Research Fund and National Science Foundation Awards CDA-9401021 and IRI-9409263.

References

[Callan et al. 1995] James Callan, Zhihong Lu, Bruce Croft (1995). "Searching Distributed Collections With Inference Networks", 18th Annual International ACM SIGIR Conference on Research and

Development in Information Retrieval, 1995

[Voorhees et al. 1994] E. M. Voorhees, N. K. Gupta, and B. Johnson-Laird, "The Collection Fusion Problem", in The Third Text REtrieval Conference (TREC-3), NIST special publication 500-225 (D. K. Harman, ed.)

[Balabanovic et al. 1995] M. Balabanovic, Y. Shoham, Y. Yun (1995). "An Adaptive Agent for Automated Web Browsing", Journal of Image Representation and Visual Communication 6(4), December 1995

[Knoblock et al. 1994] A. Knoblock, Y. Arens, C. Hsu (1994). "Cooperating Agents for Information Retrieval", Proceedings of the second international conference on cooperative information systems, University of Toronto Press, Toronto, Canada, 1994

[Arens et al. 1993] Y. Arens, C. Chee, C. Hsu, C. Knoblock (1993). "Retrieving and Integrating Data From Multiple Information Sources", Journal on Intelligent and Cooperative Information Systems, 2(2), 1993, Page 127-158

[Selberg et al. 1995] Erik Selberg, Oren Etzioni (1995). "Multi-Service Search and Comparison Using the MetaCrawler", WWW4 conference, December 1995

[MetaCrawler] MetaCrawler search home page, URL: <http://www.cs.washington.edu:8080/>

[Dreilinger] Daniel Dreilinger, Savvy Search Home Page,

URL: <http://www.cs.colostate.edu/~dreiling/smartform.html>

[ProFusion] ProFusion search home page, URL: <http://www.designlab.ukans.edu/ProFusion.html>

[Sun] Sun Microsystems, Inc., Multithreaded Query Page, URL: <http://www.sun.com/cgi-bin/show?search/mtquery/index.body>

[Cross] William Cross, All-in-one Search Page, URL: <http://www.albany.net/allinone/>

[Alta Vista] Digital Equipment Corporation, Alta Vista Home Page, URL: <http://altavista.digital.com/>

[Excite] Excite home page, URL: <http://www.excite.com/>

[InfoSeek] InfoSeek Corporation, InfoSeek Home Page, URL: <http://www.infoseek.com/>

[Lycos] Lycos Inc., Lycos Home Page, URL: <http://www.lycos.com/>

[Open Text] Open Text, Inc., Open Text Web Index Home Page,

URL: <http://www.opentext.com/omw/f-omw.html>

[WebCrawler] WebCrawler home page, URL: <http://www.webcrawler.com/>

Managing Interdisciplinary Project Teams Through the Web

Robin E. Goodman, Graduate Student, School of Civil and Environmental Engineering, Construction Engineering and Management Dept., Georgia Tech, Atlanta, Georgia, USA, robin@eiffel.ce.gatech.edu

Paul S. Chinowsky, Assistant Professor, School of Civil and Environmental Engineering, Construction Engineering and Management Dept., Georgia Tech, Atlanta, Georgia, USA, pchinow@ce.gatech.edu

Abstract: The introduction of new communication technologies such as the World Wide Web are creating unique opportunities for Architecture/Engineering/Construction project teams to develop new coordination and communication strategies. Of particular interest is the capability of teams to interact remotely in a virtual team environment. However, this evolution of project team interactions is introducing a diverse range of new issues in project management and process control, requiring a new generation of management understanding. This paper introduces these managerial issues as developed through studies of interdisciplinary, virtual project teams working together over the World Wide Web.

INTRODUCTION

Project teams in the Architecture/Engineering/Construction (AEC) industry are unique entities, created through a complex integration of factors. These entities are comprised of a network of interdisciplinary players, with varied roles, responsibilities, goals, and objectives. They are created to produce complicated project solutions through the sharing of highly specialized knowledge. However, the manner in which AEC project teams interact is rapidly changing. Project teams are increasingly becoming virtual teams; groups of individuals formed on the basis of global collaborative efforts, using information exchange and technology as a common thread to bind themselves together. This collaboration methodology is a fundamental difference between AEC virtual project teams and previous AEC project teams. The evolution is transforming how project players communicate, collaborate, and cooperate in their project undertakings. At the core of this change is the manner in which teams are exchanging and sharing information. In response to this transformation, new approaches to project management must be developed which address both the managerial and organizational issues associated with virtual project teams [Handy 1995].

One approach to this evolving project management requirement is currently being examined by the Construction Engineering and Management program at Georgia Tech. In this effort, an integrated, quality-based managerial framework is being developed to facilitate the remote communication and coordination requirements of interdisciplinary project teams. This paper introduces the managerial framework research and its unique use of the World Wide Web as a communication device to study interdisciplinary project team coordination within the classroom.

FRAMEWORK

The interdisciplinary project team research explores the dynamic interactions that affect the decision-making processes of project participants in virtual AEC project teams. The premise being that although there is a great deal of theory and practical effort directed toward understanding the managerial and organizational dynamics of project teams, a fundamental paradigm shift is now occurring toward increased use of new technologies within the project process. To understand this shift, the evolving project processes must be studied in the context of how they impact project team interaction and the resulting decisions reached by the project participants. Similarly, a new management framework emphasizing new project communication and coordination techniques must address two primary elements: 1) the dynamics associated with the remote interaction of project players during decision making processes, and 2) the new organizational structures required for the planning phases and activities of AEC projects.

Each of the management framework issues is directly dependent on understanding the unique interactions associated with virtual teams. Given this interdependency, the virtual team research effort is initially focusing upon a study of the impact of technology on interdisciplinary interaction within a project circumstance. Specifically, the study is emphasizing the need to understand how the introduction of communication tools is impacting the manner in which virtual team participants communicate and coordinate throughout a project life-cycle.

CLASSROOM INVESTIGATION

To facilitate the virtual team study, a controlled environment was required to minimize the number of external, non-controlled variables. The environment selected is the graduate level construction management course within the Construction Engineering and Management program at Georgia Tech. The course contains 43 students from a variety of professional backgrounds including construction management, civil engineering, and architecture. This diversity of backgrounds provides the essential element of interdisciplinary cooperation required to conduct the research experiments.

The use of the classroom environment provides the research team with the advantages of controlling both the information that the students receive and the diversity of the project teams. Given this control, the research team has undertaken the exploration of the dynamics associated with coordinating virtual project teams in their development of a bid proposal. Specifically, the focus of this first phase research effort centers on the effects of utilizing the Internet and the World Wide Web to share information among project participants. Through a combination of World Wide Web technologies and traditional project interactions, the research team has created a unique environment in which to observe the impact of new communications technologies. The results achieved from the study are providing a baseline from which to expand the research effort into the domain of professional project teams.

PROJECT SCENARIO

The students were provided with an assignment to complete the design, create the schedule, and provide a cost estimate for a 20,000 square foot educational facility. The classroom simulation required the expertise of architects, engineers, and construction managers within each team. Two consortia were formed, each comprised of five interdisciplinary teams-foundation, concrete, masonry, roofing, and finishes-responding to specific components of the project. Each team consists of 4-5 students with diverse backgrounds and experiences. The five teams in each consortium must collaborate to develop an organizational framework, work breakdown structures (WBS), and provide cost and schedule estimates for their bid proposal. The critical element of the project is the focus on communication and coordination. The division of responsibilities between the five teams within each consortium requires the teams to cooperate throughout the project in order to complete the overall tasks. Two central themes face the teams as the basis for the project: 1) the establishment of management mechanisms to organize and coordinate project players, and 2) the development of technical solutions to the stated issues.

The two issues are closely linked in that the successful generation of technical solutions is directly dependent on the establishment of coordination and communication mechanisms between the teams. While the students are presented with both issues as tasks which must be completed, the connection between the tasks is not provided. It is left to the teams to make this link and understand the implications of successfully setting up the coordination structures. To facilitate the study of this coordination process, each team is required to retain a log of every meeting conducted throughout the project.

PROJECT COMMUNICATIONS While initial research has demonstrated that the Web can be an effective medium for one-way communication of information from professor to student within a course environment [El Kordy 1994], the effectiveness of the Web as a two-way communication mechanism between interdisciplinary teams is less apparent. Although many factors and issues including familiarity with technology, age of participants, location of team members, and access to technology will combine to influence this effectiveness, initial emphasis must be placed on "when" and "how" the mechanism is used in actual project situations.To facilitate the coordination and communication study, a Web site was established for each consortium. The site provided two essential elements for each consortium; 1) access to individual pages containing project information required by each team, and 2) a communications center where electronic mail messages could be sent to any team within the consortium, or to the entire consortium. Each student in the class was given access to the site and given instructions on how to use the communications center within the site to ensure that knowledge and access could be eliminated as external factors influencing the study. To ensure that the study could accurately gauge the impact of the Web as a communication device and as a coordination facilitator, the students were given the option of using the Web communications center or conducting face-to-face meetings. This option was provided to ensure that the Web was being used as a voluntary communications medium and not as a forced requirement which could have corrupted the study results. Following this introduction, the students were given the assignment to develop company and consortium management structures together with work breakdown schedules for the facility.

INITIAL RESULTS

The introduction of the Web into the construction management course represents a transition step within the overall study of the Web as a communications and coordination tool for interdisciplinary project teams. For comparison purposes, the same project was given to last year's class, but without the aid of the Web as a communications tool. Similarly, next year's class will be provided the same project with the requirement that all communications must occur through use of the Web. The combination of these three stages will provide a complete comparison of the impact of the tool within the project teams. Similarly, the study will be expanded to include professional teams to compare the results obtained within the classroom to those obtained within the professional domain. However, at this time, the project team has completed the first two stages of the project and has obtained sufficient data to summarize several observations and trends.

TOTAL COMMUNICATIONS

The first step in analyzing the effectiveness of the Web in the overall communications and coordination process is to analyze the number of communications which occurred during the project. [Fig. 1] provides a summary of this information for both of the consortia in this year's class and the total from a representative consortium from last year's class. In this graph several pieces of information emerge as interesting data points. First, a comparison of this year's consortia meetings indicates a distinct difference between the number of face-to-face meetings held and the number of meetings conducted electronically. Second, the fact that the number of face-to-face meetings held by each consortium in this year's class were greater than the number of communications conducted through the Web indicates that the security of face-to-face meetings is still a strong communications consideration for project teams. Finally, the overall number of meetings between last year's teams and this year's teams is significantly increased. In each case, there was over a 66% increase in communications among the teams in this year's consortia. This increase leads to the second area of analysis, the focus of the project communications and meetings.

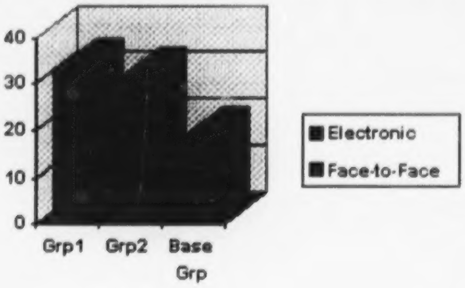


Figure 1: Total communications by study groups.

COMMUNICATION FOCAL AREAS The number of meetings called in any project can lead to one of two results-excessive amounts of time lost due to unproductive meetings, or the reduction of project-related unknowns. The result obtained for any specific project is highly dependent on the leadership of the team and the cooperation obtained between team members. One indicator which may be used to determine this level of cooperation, and ultimate focus on critical project issues, is the number of times a project group focuses on management issues versus technical issues. [Fig. 2] provides an overview of this measurement from the three focal groups in this study. Whereas last year's group had an almost 50% split in terms of management versus technical issues, this year's groups dropped to 44% and 31% in the number of focus discussions on technical issues. This significant decrease in technical focus must be examined to determine the answer one of two questions: 1) Is the decrease a result of increasing communications on the roles and responsibilities within the consortia; or 2) is the decrease an indication of greater friction within the teams which is surfacing through electronic mail messages? This question sets the stage for the next phase data analysis.

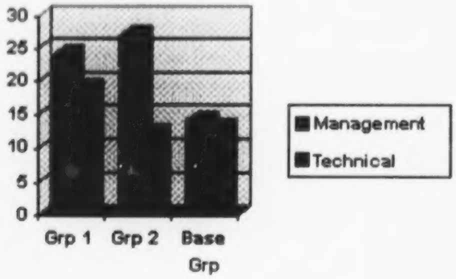


Figure 2: Comparison of Management and Technical Communications

TIMELINE FOCUS

The final area of focus is a timeline analysis resulting from the management and technical discussions. In an optimum situation, teams should rapidly decide on a management organization and then focus the majority of their efforts on solving technical issues [Bounds, et al. 1994]. However, the personalities and individual goals evident within every interdisciplinary team cannot be overlooked [Katzenbach and Smith 1993]. These external issues constantly create friction within project teams [Krackhardt and Hanson 1993]. The research team established an hypothesis that the introduction of the Web communications center would reduce the time required to focus on management issues. [Fig. 3] illustrates the actual results from the timeline study. As the figure illustrates, the hypothesis did not materialize. In reality, the management discussions continued throughout the project while the technical discussions dropped off well before the management discussions ended. This discrepancy represents a notable divergence from the initial project expectations and sets the stage for critical follow-up study.

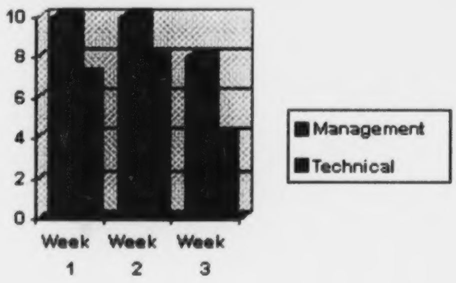


Figure 3: Analysis of management vs. technical discussions in virtual team interactions.

ANALYSIS

The introduction of new communication and coordination technologies into the project management domain is creating unique capabilities for interdisciplinary teams to approach given problems. However, the introduction of the tools does not eliminate the need to effectively communicate and coordinate throughout the project. The initial results obtained from the current study indicate that the introduction of electronic communications does enhance the number of communications that take place within a project. However, the results also indicate that the issue of coordination does not subside with the introduction of the Web capability. Rather, the Web provides the teams with the ability to continue discussions throughout the project without having to meet on a face-to-face basis. The issue thus arises as to whether or not the introduction of virtual communications exacerbated friction within the team.

In addition to the primary issue of coordination impact, the initial results from the Web study have provided several indicators for additional analysis investigation. Specifically, traditional methods for coordinating project players, sharing information, and managing teams need to be measured against the interactions of the virtual approach on the basis of the following criteria:

- clarity and mutual understanding of project requirements;
- types of information exchanged;
- efficiency of project developments; and
- quality of results.

These comparative assessments are beneficial to the continued development of the management framework, providing an ability to evaluate the cooperation and collaboration among AEC project participants. These findings will clarify the extent to which information is exchanged and effectively utilized within an interdisciplinary, virtual environment.

CONCLUSION

The managerial framework being developed through this research effort will become an innovative instrument for the AEC industry. By incorporating modern management theory with the assessment of team dynamics on project planning activities, a comprehensive, practical model will be developed. This framework will support the organizational and managerial requirements of interdisciplinary AEC virtual project teams. Specifically, it will provide a strategy for team creation and coordination, interdisciplinary sharing, team leadership, and client understanding. Subsequently, through comparisons of traditional project interactions, the research will provide AEC project participants with a new strategy for collaboration, including a methodology for application in real team settings.

REFERENCES

Bounds, G., et al. (1994). *Beyond Total Quality Management. Toward The Emerging Paradigm.* New York, NY: McGraw-Hill.

El Kordy, O. M. (1994). *The Application of Multimedia and Knowledge-Based Systems to Computer-Aided Engineering Instruction.* MS Thesis, Nelson Baker, advisor.

Handy, C. (1995). Trust and the Virtual Organization. *Harvard Business Review*, May-June, 40-50.

Katzenbach, J.R. and D.K. Smith (1993). *The Wisdom of Teams; Creating the High Performance Organization.* Boston, MA: Harvard Business School Press.

Krackhardt, D. and J.R. Hanson (1993). Informal networks: The Company Behind The Charts. *Harvard Business Review*, July-August, 104-114.

MHEG-5 Delivery in the DAVIC Environment

Chetan Gopal , Cesar A. C. Teixeira and John F. Buford

Abstract: The Digital Audio-Visual Council (DAVIC), an international consortium, has published specifications for the components of interactive television (ITV) systems to interoperate. DAVIC 1.0 requires ISO MHEG part 5 (Multimedia Hypermedia Expert Group - Coding of Multimedia and Hypermedia information) as a supported content type at the STB, and uses MPEG2 DSM-CC (Digital Storage Media Command and Control) as the basis for the interaction protocol between the clients and servers. This paper is aimed at eliciting the relationship between DAVIC, DSMCC and MHEG and how ITV can be integrated with the WWW.

Introduction

The WWW provides a means for users to access media, both discrete (DM) such as text, images, and data and continuous media (CM) such as video and audio, from various services over the Internet. Today these media are embedded in HTML documents without synchronization or real-time delivery. There are no quality of service (QOS) negotiations between the client and the server. The servers are generally stateless and unaware of client profiles to make such QOS decisions. Several companies including Netscape, Xing Technologies, and VDOnet extend the *httpd* server by adding a separate streaming media server. These servers maintain a separate connection with the client to deliver CM and this connection may involve transport using a special QOS-aware protocol, for example, RTP (Real-Time Protocol). In the future it is expected that the introduction of RSVP (resource reservation setup protocol) into the Internet will make it feasible to support large scale CM use in the Internet. In the meantime, existing solutions are proprietary and seem suited primarily to intranet deployment. Further, the HTML document model has no notion of time composition.

The ITV world, generally viewed as a joint activity of computer, entertainment, cable, and phone companies, has been working on a system specification that would make the main components of an ITV system interoperable. These components include the Set Top Box (STB), the network service providers, and the content providers. The first version of this specification, DAVIC 1.0, was released in December 1995, and incorporates a number of existing standards including MPEG-2 systems, MPEG-2 DSM-CC, and MHEG-5. DAVIC is currently working on a new version of the specification that will include support for accessing Internet and WWW services, including the ability for the STB to execute Java applets. It is expected that the Java VM will be integrated with the MHEG-5 engine to permit Java and MHEG content to be combined; this integration activity is termed MHEG-6. A DAVIC 1.0 interoperability trial has been successfully conducted in June 1996 [Columbia 96].

ITV systems will use a low-cost Set Top Box (STB) [Furht et al. 95] that can be configured remotely to connect to the outside world. They will also use new high bandwidth network architectures like ATM, HFC, and ADSL to provide the necessary higher bandwidth to the consumer. Most ITV applications will be downloaded incrementally because of memory limitations. It seems likely that in the future many consumers will have access to high-bandwidth information services through an ITV provider such as a cable TV or telephone service. Consequently, the ITV standards developed by DAVIC are likely to be important for developing web client and server standards which support real-time video and audio delivery. Similarly, MHEG-5, which provides both a time-composition model and an object-oriented framework for incremental delivery of content, may also be relevant to future extensions to HTML, which currently lacks these key features.

The purpose of this paper is to briefly review the DAVIC model and MHEG-5 with the expectation that there will be a future relationship between deployed ITV systems and Internet and web systems. Table 1 compares features of the WWW and DAVIC.

Features	WWW	DAVIC
Document Encoding	HTML	MHEG-5 required, other formats optional
Real-time delivery	No	Yes
Stream interface	None	DSM-CC Stream functionality
Session Management	None	DSM-CC (U-N protocols)
Inter Media Synchronization	None	Through MHEG-5 links

Table 1: Features of WWW and DAVIC

Overview of MHEG-5

MHEG-5 [MHEG 95] belongs to the suite of standards known as MHEG [Buford 94]. It defines the semantics and final form for interchangeable objects intended for composition of simple interactive multimedia applications in a client/server architecture, ensuring the interoperability across heterogeneous platforms. MHEG-5 like its predecessor MHEG-1, is intended for low-end delivery systems. MHEG-5 has a more precise conformance requirement than MHEG-1, and is generally considered to have a clearer and less complex design.

Multimedia applications in the context of MHEG-5 are event-driven. They are composed of a web of scenes and link objects which define navigating between scenes and the actions that must be executed when some specified events occur. Most of the applications can be developed using only declarative code, although MHEG-5 provides means to perform tasks which are better expressed in a procedural paradigm.

A scene object groups visual, audible, interaction, and link objects. Consequently, the application and all its scenes with related objects, are designed only once and delivered using MHEG-5 object Servers. The objects are downloaded to the client and the MHEG-5 run-time engine is responsible for decoding, interpreting, presenting, and managing them.

Design and Implementation of an MHEG-5 Engine

This section describes the design and implementation work which is being carried out by the authors. The process or set of processes that interpret MHEG-5 objects and deliver them to an application for presentation is referred to in the standard as an MHEG-5 engine. A link engine processes such events and associated data, and fires the links that have their link condition satisfied. The model adopted by the authors is shown in Figure 1.

The Object Processor (OP) is the module responsible for the effective interpretation of MHEG-5 objects and execution of actions. When Link objects have to be activated, the OP registers them in the Link Engine (LE), where a table mapping Link-Conditions into Link-Effect is kept. The Object Store (OS) supplies prepared objects to the OP, and the Object Decoder (OD) deals with decoding and preparing MHEG-5 objects.

When the MHEG-5 engine is first activated, it runs a cold start process to retrieve the first Application object to be processed. From this first object users can choice their desired Application object.

Once the OP gets the identification of the object to be processed, it requests the prepared object from the OS module. The OS keeps a cache with Scene objects prepared in advance, with the aim of minimize access time, and objects that had been active before and have high probability to become active again in a short time. If the OS does not have the required object, it requests the OD to decode the object and then prepares it.

Based on the type of the object, the OP knows what attributes to look for in the data structure resulting from the object preparation. If it is processing an Application or Scene object then it starts the execution of the Action object specified in their On-Start-Up attribute. All elementary actions are queued and processed accordingly. Link objects found by OP in Scene or Application objects are activated and registered at the Link Table.

Some events may occur as a consequence of the OP processing activities (synchronous events) or generated by processes that are asynchronous to the MHEG-5 engine, e.g. user interactions, firing of the timer, etc. In the later case, the occurrence of events are passed on to the OP through the primitive External Input(). These external events are translated to proper events recognized by the standard and, same as happen with the synchronous events, they are reported to the LE module.

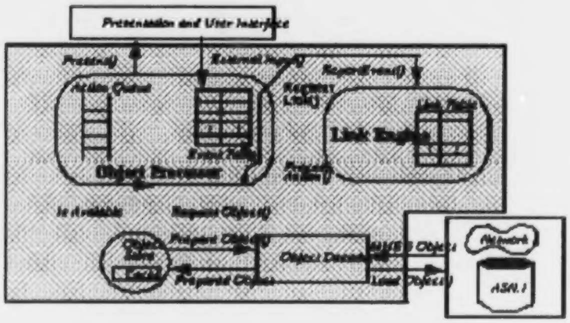


Figure 1: MHEG-5 engine model

On receiving the information that an event has occurred, the LE checks active Links of all active Application and Scene objects to determine whether some Link must be fired or not. When a Link is fired, the LE uses the primitive Process-Action() to inform the OP which Action object should be processed.

Overview of DSM-CC

The DSM-CC is a set of protocol specifications for managing and controlling MPEG-1 and MPEG-2 bit streams. These protocols are intended for general applications, MHEG applications, and scripting language application to setup or teardown a network connection using User-Network (U-N) primitives and for communication between a client and a server across a network using User-User (U-U) primitives. U-N primitives are defined as a series of messages to be exchanged among the client, network and server. U-U primitives may use a Remote Procedure Call (RPC) protocol and may involve U-N messaging. DSM-CC may be carried as a stream within an MPEG 1 System Stream, an MPEG 2 Transport Stream, or an MPEG 2 Program Stream. Alternatively, DSM-CC may be carried over other delivery mechanisms, such as TCP or UDP.

Overview of DAVIC

The DAVIC is a consortium whose purpose is to favor and nurture emerging audio-visual standards and services. These goals are achieved open international collaboration of all members of the DAVIC consortium. As of today, there are more than 200 companies from 20 different countries who participate in DAVIC activities. The DAVIC specification is intended to be open and to specify tools, protocols for handling information flows and not systems. Thus different vendors can make their systems interoperable if they conform to the DAVIC specification. DAVIC specifies minimum requirements only and it is up to the vendor to add more functionality.

DAVIC Architecture

The DAVIC system specification is aimed at providing interoperability. It is based on an abstract reference model and is made up of 5 subsystems (see [Fig. 2]). A11, A10, A9 and A1 denote reference points within the system and indicate the information flow between modules. A reference point comprises of one or more logical information-transfer interface and one or more physical information transfer interface. DAVIC 1.0 system typically addresses the service provider system (SPS), service consumer system (SCS) and the delivery between the two. The DAVIC system uses a four-layer service model, namely, principle service model(SL1), application service layer (SL1), session and transport service layer (SL2) and network service layer (SL3).

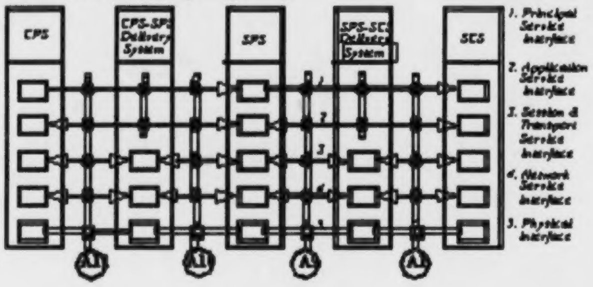

CPS
CPS-SPS Delivery System
SPS
SPS-SCS Delivery System
SCS
1. Principal Service Interface
2. Application Service Interface
3. Session & Transport Service Interface
4. Network Service Interface
5. Physical Interface
1
2
3
4
5
A11
A11
A
A

Figure 2: DAVIC System Architecture (source: [DAVIC 95])

The DAVIC protocol specification is based on the DSM-CC specification. All U-U (User to User) signaling from the STU to service provider takes place over RPC, and content and data download is carried by MPEG-2 transport packets. To pass a message the STU will need to do a number of processing including message fragmentation and protocol encapsulation. These protocols are classified into the following information flows, namely, S1 flow for content and data delivery, S2 flow for U-U messaging and S3 flow for U-N messaging.

Content Provider System (CPS)

The main function of the CPS is to provide for functionality for content providers to interact with the SPS for loading and updating content information at the SPS using the CPS-SPS delivery system. The current specification of DAVIC 1.0 does not deal with either CPS or CPS-SPS delivery system.

Service Provider System (SPS)

This system provides access for consumers to various services in the system. It comprises the 1) Content Service Element (CSE) which provides processing for the SPS when content information is being loaded to the server, 2) Application Service Element (ASE) which manages and processes application control flow information, 3) Stream Service Element (SSE) which provides for processing of content information flow at the stream level, 4) Session Control Function (SCF) which provides processing for establishing & terminating sessions and specifies quality of service (QOS) requirements for both the application and product entities, 5) Network Control Function (NCF) which provides for error-free transmission and reception of content information flow to and from the server, and 6) Network Interface Function (NIF) which handles network specific information flows and provides a non-network specific interface to the rest of the SPS.

Service Consumer System (SCS)

SCS comprises of the Set Top Unit(STU) or Set Top Box (STB) - This unit helps the human or machine peripheral system interact with service providers. The STB comprises of the following components, 1) Product Entity - This is responsible for accepting content information and presenting it to the peripheral system transparently, 2) Application entity - Will handle interactive information processing for the consumer, 3) Environment entity - Responsible for establishing and terminating the environment in which an application will operate. QOS requirements for both consumer and product entities are a part of this environment, 4) Connectivity entity - handles error-free transmission and reception of content to and from STB, 5) Network interface unit - Provides a gateway for STB to communicate with the delivery system. It is through this unit all content information flow will take place. It also provides for a non-network specific interface.

SPS-SCS delivery system

This system defines a model for transferring audio and video information from service provider system to service consumer system. This system consists of 1) Service related control - Supports for SL0, SL1 and SL2 related control functionality, 2) Network related control - Supports for network configuration (SL3), connection setup and termination, and information routing, 3) Core network - represents a high speed digital network and provides a means for accepting and transmitting information flow in an error free manner, and 4) Access network - This comprises of: a) access node, which is responsible for processing information flow for transport through a distribution network, b) Distribution network, which provides for transport medium for delivering information from source to destination, and c) Network termination, serves as a bridge for the system for adaptation of information flow from one network configuration to another.

Scope of MHEG-5 in DAVIC

DAVIC 1.0 requires MHEG-5 as a run-time environment (RTE) for STB, it does not require that all implementations of MHEG-5 for STBs provide for complete MHEG-5 implementations. It identifies all required MHEG-5 specifications. In addition to the above, DAVIC requires RTE to use a standard set of user input events, connection management, standard set of RPC interfaces, persistent name spaces and standard encoding of content data. In addition this, DAVIC defines the MHEG-5 to DSM-CC U-U mapping in terms of name-space mapping, high-level API for object and content references. By defining all these requirements conformance is achieved. Using MHEG-5 on the STB provides for the highest degree of platform independence and interoperability.

Integration with the WWW

MHEG-5 supports text objects encoded using HTML, and also supports URL-based addressing of objects. When ever an HTML content is referenced through an MHEG-5 hypertext object, the corresponding WWW service provider is identified and the object is retrieved from the web, packaged as a new scene with hypertext ingredient and delivered to the MHEG-5 engine. The engine does some processing on this new scene object and then presents it to the user. Figure 3 shows how a DAVIC STB system can interact with a WWW server to gain access to the WWW.

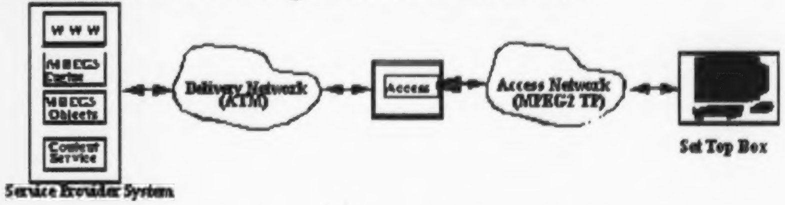


Figure 3: DAVIC-STB used in a gateway to WWW

Figure 4 shows the DSM-CC message flows. Here the STU uses the U-U interface to communicate with the WWW server which uses the DSM-CC download protocol to transfer HTML content The scenario shown here is flow controlled and it is through the U-U messaging the STU and server decide upon window size for download and other download requirements before using the download protocol to retrieve the HTML object.

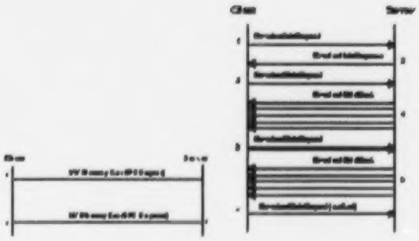


Figure 4: DSM-CC message flows for downloading HTML content using download protocol.

Conclusions and Future Work

We see the future of ITV as not just limited to VOD but to a more general service for users to access a variety of information. In this context, DAVIC clearly specifies an architecture for ITV that can be effectively deployed and existing services be easily accessed.

We are in the process of developing an MHEG-5 engine and simulating the MHEG-5 latency budgets for DAVIC ADSL networks.

References

[Buford 95] Buford, J., Gopal, C. Standardizing a Multimedia Interchange Format: A Comparison of OMFI and MHEG, Proc. International Conference on Multimedia Computing and Systems. May 1994, Boston, MA.

[Columbia 96] Columbia University. Proc. Workshop on Video on Demand Systems: Technology, Interoperability, and Trials. June 24, 1996.

[DAVIC 95] DAVIC, DAVIC 1.0 Specification Parts 1-12, Revision 5.0, Nov 16, 1995.

[DSMCC 95] ISO/IEC, DSMCC: Digital Storage Media Command & Control, ISO/IEC 13818-6 DIS, July 1996.

[Furht et al. 95] Furht, B., et al. "Design Issues for Interactive Television Systems", IEEE Computer, May 1995, pp. 25-39.

[Koegel 92] Koegel Buford J., On the Design of Multimedia Interchange Formats, Proc. Third international Workshop on Network and Operating System Support for Audio and Video, Nov. 1992.

[MHEG 95] ISO/IEC. MHEG: Base Object Encoding. ISO/IEC 13522-5 DIS. December, 1995.

Acknowledgments

Cesar Teixeira is supported by CNPq/ProTeM-CC/SMmD, Project number 680077/94-4

Aster Servers - How to make Aster more widely Available

A. Hashmi, B. Barry, R. Lundquist, and M. Krishnamoorthy
Bellcore, Morristown, NJ.
Science Access Project,
Oregon State University, Corvallis, Oregon.
Department of Computer Science,
Rensselaer Polytechnic Institute, Troy, NY.

This paper describes two client-server models for using the audio formatting system ASTER. ASTER was developed by T.V. Raman[1] at Cornell University to produce high quality speech output from documents marked up in TeX or LaTeX. ASTER is written in Lisp, and the normal way to run it is interactively through an emacs interface. However, in order to provide a more universal way for people to use ASTER, we have developed two client-server interfaces. The first allows a user to submit marked up documents in an html form to ASTER running on the server and to receive back speech commands which can be redirected to the client's speech synthesizer. The other interface allows a person to run ASTER on a server through a telnet session but to redirect the synthesizer commands back to the client machine.

Introduction

ASTER (the Audio System for Technical Reading) takes documents written in a markup language such as TeX or LaTeX and provides audio-formatted speech output through a DECtalk speech synthesizer such as the Multivoice or the DECtalk Express. The history of ASTER is that it was originally written in Lucid Common Lisp and was then ported to CLISP. It currently runs on a variety of workstations including Sparcs, HPs, DEC Alphas and PCs running Linux. The usual way to run ASTER is interactively through an emacs interface. This emacs interface allows the reader to put the marked up document in an emacs buffer and submit it to ASTER, which is running as a Lisp subprocess in another emacs buffer. The user is then able to browse the document using key sequences bound to emacs Lisp commands.

This way of using ASTER is fine if the user has available a unix machine running Lisp, but it is not a general solution for a large number of users. However, since many users do have World Wide Web access using browsers such as W3, Lynx and Netscape, we have therefore developed an html-based client-server system for using ASTER.

Web access to ASTER

The Web client-server interface to ASTER is based on an html form. The user puts his marked up document into a multiline form space on an html form and submits that form to the server, which then runs ASTER on the submitted document. The output of ASTER is a string of synthesizer command sequences which are returned to the client and routed to the hardware speech synthesizer.

Client

The client is any system running a Web browser which has forms capability, since an important feature of this client-server system is the creation of a "form" in html. This form is filled in by the user and submitted back to the server. The form type used does not limit the length of the input, which may be a single sentence or an entire book. The user has two options. Either he can type the data directly into the form area or he can "paste" it from another window.

Server

The server is responsible for receiving the html form containing the document, parsing it, passing the document to ASTER and then returning ASTER's output back to the client. The server we are using is an httpd server installed on the Rensselaer computer science network. It is running on "Juicer", one of the more powerful machines at RPI (a Sparc 20 with 256 MBytes of main memory and 503 MBytes of virtual memory). This httpd server provides ready access to clients connected to the World Wide Web through browser applications that use the http protocol, e.g., Netscape and W3. This server allows the system administrator to set different permissions for different users and to set restrictions such as bandwidth limitations. Most importantly, this server has the capability to execute scripts on the local machine.

The httpd server is configured to handle scripts written in BASH as well as some executables written in Lisp, C and Perl. BASH is used because it has better program control transfer capabilities and because some system variables can be defined internally in BASH, whereas they must be defined externally in some other shells. For security, the server can only execute scripts kept in a special directory. Any paths and links used by these scripts must be given relative to this directory. This is important because several programs have to be called in order to execute ASTER.

The first script takes the incoming html form and passes it to a parser written in Perl. The purpose of this parser is to strip out of the form just the relevant document. The parser reads the contents of the form, checks for any errors, and if all goes well, writes the TeX document to a file called "/tmp/temp-msk".

The next stage in the script sequence is to execute ASTER. ASTER is run directly from a memory image without the emacs interface. Usually the emacs interface is used to pass the document to ASTER. However, we used a function of ASTER which retrieves the document directly from a file so that it can be run without emacs. The input file is "/tmp/temp-msk" which was created by the previous parsing step. ASTER takes this input TeX file and returns a set of DECtalk commands that can be sent to the voice synthesizer to produce speech output. Normally DECtalk commands are mixtures of the words to be spoken along with commands to change pitch, speaking rate and other voice parameters. These commands are saved in another file.

The third stage in the script sequence is to return the output DECtalk commands to the client. To do this, we define a new MIME type for ASTER. This MIME type allows the client to accept the DECtalk commands and redirect them via the serial port to the hardware voice synthesizer.

Since the Lisp processes take some time to complete execution, depending on the document length and complexity and the server load, it is necessary to advise the client that his/her request is being processed. For the sighted user under X-Windows, we created a special widget which "pops-up" whenever the client sends a request to be processed and advises the user that his request is being processed.

A special "lock" mechanism has also been incorporated into the server. This allows the system administrator to "lock-up" the server when it is in use so that many people will not be using it at the same time. This is needed because the data input and output files that are created would get overwritten if another user input his file. The server can easily be modified to allow several users to access ASTER at once. In this case, the input data files would simply be given a unique filename based, for example, on the pid of the individual process handling a specific user. However, this has not been done in the current setup, since our server is not running on a dedicated computer and this would take up too much memory and CPU processing time. Running multiple Lisp sessions will only be suitable when a machine is dedicated entirely as an ASTER server. The other possibility is to have a single ASTER running that takes the input from multiple users simultaneously. We are currently working on this design.

Also it would be useful to maintain on the server a persistent data-base of "books" that the client could read. The user would then select from a list of books to read or browse. These books, which would have already been processed by ASTER, could be quickly returned to the client, enabling faster response times for the client and lower overhead for the server.

The Web based server has been tested by users from DEC Cambridge Research Laboratory, Recordings for the Blind, Rensselaer Polytechnic Institute and Oregon State University. The main drawback is the long response time, which is partly due to the load on the network and partly due to the load on the server.

Telnet Interface

Another approach we have tested as an ASTER client-server interface uses telnet. The server is a 486 running Linux and the client is a 486 running Windows. The user on the Windows client logs onto the Linux machine using telnet and then runs ASTER using the normal ASTER-emacs interface. Normally ASTER sends the DECtalk commands via the serial port to the hardware voice synthesizer. However this is useful only if the user has direct access to the Linux machine. Instead, to get the DECtalk commands to the client, we send them via a pipe to another process running on the server. This process then takes the DECtalk commands and sends them through a Berkeley Socket over the internet to a process running on the Windows client. The Windows client then sends the DECtalk commands to the synthesizer via the serial port on the client and also sends any responses from the synthesizer back through the socket over the internet to the server. We chose a Windows client because Windows is widely available and because, using Winsock, the client can run both telnet and the process which redirects the commands from the socket to the serial port.

Notice that this method requires at least two processes on the server and two processes on the client. The client is running both telnet and the socket to serial port redirection process; the server is running both ASTER and the serial port to socket redirection process. This trick of redirecting the DECtalk commands through an internet socket allows the user to use ASTER to read and browse through his document just as if he were directly on the server. Note that DECtalk commands, not audio files, are being passed back and forth from the client to the server. The bandwith needs are therefore quite small, and this type of interaction can be sustained over a relatively slow internet connection. The main drawback is that to use telnet the user is required to have an account on the server, unlike the Web method which allows anyone access.

Summary

Two different systems have been developed for running ASTER via a client-server interface - one an html/Web based approach and the other an internet socket/telnet approach. The Web system would make ASTER more generally available to the public while the telnet/socket system would permit more interactive use. Research is being planned to merge these systems to create a single client-server method for using ASTER. This method could then be made available to the public.

References

[1] Raman, T.V. , "AsTeR - A system for reading Technical Documents," Cornell University, 1995.

http://www.cs.cornell.edu/Info/People/raman/raman.html

Collaborative Learning and Knowledge-Construction Through a Knowledge-Based WWW Authoring Tool

Erik Haugsjaa
Center for Knowledge Communication, Computer Science
University of Massachusetts, Amherst, MA, 01003 USA
ehaugsja@cs.umass.edu, www.cs.umass.edu/~ehaugsja/

Abstract: The World Wide Web has huge potential as a medium for learning, both in and out of classrooms. As more and more schools obtain access to the internet, learners (including parents, teachers, and students) are realizing that the web is not particularly useful for substantial educational use in it's raw form. This is primarily due to the lack of collaborative knowledge-construction facilities integrated into the web framework as well as the lack of guiding structure and dynamic management of link and nodes. This paper outlines the hurdles one faces in using the web for learning, specifically in a collaborative knowledge-construction environment. It then suggests some theoretical solutions based directly on existing web environments, as well as on research and system prototypes in the areas of Intelligent Tutoring Systems (ITS) and ITS authoring systems.

Introduction

Education for the 21st century -- Global issues grounded in the local community

"The more things change, the more things stay the same." In a February 1996 interview with Wired magazine, Steve Jobs commented that when one looks at the big picture, computers haven't changed the fundamentals of what life is about, "We're born, we live for a brief instant die. It's been happening for a long time. Technology is not changing it much, if at all." [Wolf 1996] This is a powerful statement coming from Jobs, who, in a previous incarnation as the co-founder of Apple computer, made it his mission to revolutionize education by bringing Apple computers to classrooms around the United States. Regardless of Jobs' credentials, the comment does ring true, and we see it in our school systems as we struggle to make use of the exciting new world of the World Wide Web. The computer and communication technologies in general have changed things in the world of *work*, but we still struggle to find an important role for them in improving education.

Information technologies -- writing, the printing press, and onward... -- have brought a never-ending increase in the amount of information available to the general world population. And while this flow of information has proven to be useful in the workplace to those that have already gained the skills needed to filter through this infortmation, they have not, with the exception of writing, improved the overall level of education of the population. This is not to say that artifacts such as television and computers can't be useful as educational tools, it is just that currently, there negative effects far outweigh their positive effects [Mander 1991]. Also, many argue that the "medium is the message" [McLuhan 1964] and that television is doomed to have also have a net negative impact on society [Postman 1992]. However, with the web, the situation is more hopeful. The World Wide Web has increased the amount of information a person has easy access to, but it has done nothing to aid the individual in dealing with this information. In fact, like television, it often inflicts the individual with a feeling of passive submission. Informed, but inactive. This is evidenced by the small number of personal home pages and the even smaller number of those with any content other than links.

The web was originally designed by Tim Berners-Lee as a medium for active widespread publishing and collaboration between researchers, not as the one-to-many "tv-like" medium with the occasional list of a person's "favorite channels" that is most common today. [Berners-Lee, et al 1994] This situation probably developed for many reasons, including the difficulty of creating web pages directly using

HTML, and the fact that the server and client-side mechanisms for providing for at least some form of collaboration, group annotations, was only implemented in a limited form (personal and local-group annotiations, but not public annotations) by the NSCA Mosaic team at UIUC in 1993. [Annotations]

Instead of a new tool for providing interesting new ways for managing information and publishing on one's own, we are left with a tool that has been responsible for over-extending our abilities to make sense of the information available on the web. What we need today are tools to help us deal with this information and to help young people how to deal with this information on their own and most importantly, to focus and write down ideas of their own. This is not a skill which one comes by naturally -- one needs the assistance of people and tools to help provide structure, meaning and understanding to the flow of facts, theories, and ideas.

Is Science to Blame?

In some cases, the amount of information that we are faced with has increased simply because our understanding of the world on a scientific and social basis has become so intricate. In some areas of science, our understanding of the world has become so complex that it is increasingly difficult to relate the interelations and subtleties to a classroom of passive learners. Instead, it's necessary for the learners to take an active part in the learning process -- they need to actively construct understandings for themselves grounded in their everyday life and community. This is also very true for subjects like social studies, history, literature or any other topic that considers complex social, cultural or global issues. Being actively involved in the typical primary or secondary classroom can be difficult. 20 or 30 students is often too large a number of people to create an inclusive discussion where everyone is able to make their thoughts, ideas, and opinions heard. A computer tool can provide a medium where a person feels more able to at least initially have their voice heard. And people that feel they have a voice in an online discussion are likely to carry over their empowerment to their face-to-face interactions.

The MeTAL design is a system of WWW server- and a client-side additions that will create a useful collaborative learning tool in the classroom by adding functionality that automates the tasks that require the users to focus on the form of their WWW writings rather than their content. It also allows for adapting the viewing of this web of knowledge that individuals or groups have created according to the user's preferences. The primary focus of a young person's education should be on the thoughts and ideas of those in their place-based community. It has often been noted that students involved in distance learning projects with students or researchers in far-away places has not made them more intellectually curious or socially aware in their relations with students in their immediate environment. [Talbott 1994] Discussion and interaction with people they don't interact with on a daily basis are treated differently than those they have online with people that are "right-there". MeTAL's focus is on the "right-here" even when the topic is of global concern. See also [Global Lab].

Hurdles Towards Reaching a Learning Web

Hurdles - Some Relevant Web and Hypertext Background

In 1989, Tim Berners-Lee originally envisioned the WWW as a tool to be used by research scientists like himself in assisting in personal or group research projects. The WWW would be a powerful tool for collaboration, allowing one to easily organize and share one's research ideas and results with colleagues. As a cross-platform hypertext system, it would seamlessly integrate text and images, and allow one to easily create explicit logical links between implicitly interconnected material provided by multiple individuals. In contrast to this view of hypertext collaboration, Ted Nelson, the eclectic hypertext pioneer and visionary relates the following in comparing his and more traditional hypertext system designs, "...I just wanted to be left alone and given the equipment and basically to empower smart individuals and keep them from being dragged down by group stupidity. The amazing thing is that our designs have converged to some degree, showing, I think, the fundamental validity of this whole approach." [Nelson 1995]

This "group stupidity" that Nelson writes about I believe is due more to the current structure of educational environments, more than it is due to the human condition. Because of this, new forms of collaborative knowledge-construction that support both the individual and the group are needed. Nelson would be served just as well as a group of learners. And Nelson would be able to share his knowledge-space with others at will. While Berners-Lee was correct that the WWW has potential as a powerful tool for collaboration, there are at least two major stumbling blocks towards making this a wide-scale reality: 1) lack of group annotation capabilities, 2) reliance on screen or "media"-based authoring of hypertext content. Developing a system to overcome these hurdles is necessary to enable the elaborate and complex collaborative activities of group problem-solving on the WWW, whether the users are students in Jr High or physics researchers. [Annotations]

Hurdles - Annotations

The original design of the WWW included specifications for allowing both personal and workgroup annotations. Annotations in the WWW sense are simple comments attached to a document. They are visible to either just to the annotation author (personal) or to a defined set of people (workgroup). Personal annotations were implemented in early versions of the Mosaic WWW client software but workgroup annotations were not implemented likely because it was a substantial additional programming task and their were other tasks of higher priorities for the programmers of the WWW server software.

Lack of group annotations has meant that it is very difficult to make use of the web for substantial collaborative efforts. Single authors can make content available and create associative links to other already existing content, but collaborative creation and maintentence of content and links using the standard web interface and tools available is not really possible. From a technical implementation standpoint, there is one main issue that is not currently handled smoothly: link creation and maintainence. As mentioned above, creation of associative links from one's own content is currently possible. The essential problem is that creating links in *other's* content is not currently possible. This is the limitation of the WWW which group annotations were meant to deal with.

However, while the group annotation solution is relatively straightforward to implement, it is not a particularly attractive one because it is not flexible or general in the types of collaborative contributions or discussion it allows. The fundamental limitations of annotations are: "links" (not true links in the WWW sense) to annotations are attached to the bottom of the annotated page; annotations are not actual WWW documents. While it is theoretically possible for anntotations to contain HTML code and links to other WWW pages, the annotation is not a true WWW page and therefore cannot have additional annotations attached to *it*, and more limiting still, it cannot have associative links from a WWW page directly to it. [Annotations]

This is the level of tightly coupled functionality one would like to acheive with a distributed hypertext system meant for collaborative work. Right now, the WWW does not provide this sort of functionality in an automated fashion. Instead, to create a link from another person's page to one of your own pages, it means writing email to them personally and asking if they would "link you in". There are numerous systems that work on top of the WWW that attempt to improve on the limitations of the never-implemented Mosaic annotation concept. The best examples of fully automated systems include HyperNews [Hypernews] and the Threads program at the Hotwired web site [Threads]. The very best example of an online collaborative discussion with associative links between multiple viewpoints on a single issue (non-automated) is FEED Dialog [Feed]. We would like to be able to automate this level of functionality. Not to be forgotten in a discussion of related systems is Ted Nelson's still-a-pipe-dream Xanadu system [Xanadu] which would allow bidirectional links as well as automated notification of changes to marked pages using "sensors".

Hurdles - Media-Based Authoring

If one was to implement an automated method of annotating someone else's WWW pages by adding links to new WWW pages, one would have solved a large problem in preventing smooth automated facilitation of collaborative problem solving and discussions on the WWW. But another even more substantial issue would still be holding you back from making such a system assessible to those uninterested in the technical aspects of web publishing and HTML coding: the content and form of

WWW pages are intertwined. In other words, like virtually all computer-aided instuction systems of today, WWW pages are media-based, rather than knowledge-based [Murray & Woolf 1992]. What this means is that there is a severe limitation in the way that the pages are viewed by diverse users. Besides the ability of most WWW servers to easily let one set passwords on certain documents, and people having different settings for their WWW client software, the experience of all users will be essentially the same -- the presentation order of content and number and location of links remains unchanged. Group annotations would have solved this problem in a non-ideal way: users would see a different list of annotations at the bottom of documents according to the workgroup they are a part of. But a much more flexible solution can be envisioned.

MeTAL: A Knowledge-Based Tool for Collaborative Knowledge Construction

MeTAL (meta-tool for active learning) is a system designed to make students an active part of the learning process. Project-based learning environments (e.g. CoVis [Covis], LOGO [LEGO LOGO]) are a huge step in the right direction, but still leave out many students that don't feel they can make a personal contribution.

Project Based Learning - Global Village or Globe of Villages

"Being a socially competitive species, we naturally compare ourselves with people we see, which meant, in the ancestral environment, measuring ourselves against fellow villagers and usually finding at least one facet of life where we excel. But now we compare our lives with the fantasy lives we see on television... Our own wives and husbands, fathers and mothers, sons and daughters can seem profoundly inadequate by comparison. So we are dissatisfied with them and even more dissatisfied with ourselves." [Wright 1995]

Information continues to increase at a rapid pace, and with it, specialization in both work and play. It has become more and more difficult, as Robert Wright describes, to "find at least one facet of life where we excel." [Wright 1995] This is certainly true for students in our classrooms. It is hard to feel like you are an "expert" on a certain topic when you have direct access to Ph.Ds in the field at your fingertips. "Oh well. Why should I even lift a finger? The world is at my fingertips." What are the hidden messages of project- based learning projects that has communication with domain experts as a main tenet?

What happens when we let our students communicate at a young age with domain experts, or have them manipulate scientific data, collected in a far away place? Two things: The domain may be oversimplified and trivialized, and the students may be simply learning to be passive observers and manipulators of information rather than active participants in the learning process in their place-based community. Our classrooms need tools that bring students' minds back to their local community and back to the idea that THEIR ideas are ones that are worthy of respect. Students need to learn to trust themselves and not come to rely on the "big media" domain experts for the answers to their life and their community.

The Form Versus Content Issue, Moderators, Voting, Verbal Skills and Democracy

MeTAL takes the project-based approaches mentioned above one step further. Students' or groups' final work is not just presented, gathered up, and combined to produce a single class report that is never used or seen again. Students and teacher alike use the system in a collaborative effort and the system is designed to help students focus on their ideas and their writing, rather than on surface features like fonts and document formatting. The focus is on the process and not just the final product. Students are actively involved in the process of using MeTAL and students have an authentic reason to write -- to express their ideas.

In this sense, everyone that is using the system has the opportunity to contribute to the domain content of the system. However, some information (both content and viewpoints) will usually be somewhat fixed in the system as the use of MeTAL is typically moderated in the form of fixed, pre-existing links and nodes. Like a classroom discussion, the discussions that develop on MeTAL may be moderated by a teacher or student or they form around nodes created by a possible contributing domain expert. These node locations in the student's view of the group's information space need not be static. Because document views are generated dynamically, the overall sturcture can change along with the content. See [Object Oriented Database - Queries, Dynamic Viewing] below. Some of additional features such as voting on nodes is possible as in the Idea Futures WWW system which is "... like a corn futures market, where one can bet on the future price of corn, here one bets on the future settlement of a present scientific controversy..." [Hanson 1990]

While developing good verbal skills is an important part of one's education, a classroom environment with 30 other students wanting to speak does not aid this development. In the end, the person doing most of the talking is the teacher. This is not the case with MeTAL. There is still the possibility that students get left out of the discussion on a certain day, but it won't be because they didn't get the chance to contribute. And for these students, much is gained by having the chance to traverse through MeTAL's nodes, following the direction of the person or people of their choice, thanks to the filtering capabilities of the system. The user can create, with the help of the system, their view of the topic, filtered both by difficulty of the node concepts and author.

Students play only a secondary role in the learning process in the typical primary and secondary school today. The teacher is doing most of the talking and it is clear that the classroom learning is in the hands of the teachers. MeTAL is a tool which can be used in a democratic, student-centered learning environment. Using MeTAL, students become active participants in the learning process. MeTAL breaks down the barrier between student and teachers and allows everyone's thoughts and ideas to be expressed. [Sudbury Valley School]

Hypertext is not inherently educational. [McKendree 1995] MeTAL is useful to students in secondary schooling that are at the point where they are able to engage in meaningful discussion of topics that goes beyond mere facts. Students benefit from MeTAL because they are able to experience different viewpoints on a given topic and take part in actively seeking out these different perspectives as well as generating ideas and opinions of their own. They are able to go further in the development of their ideas in a given domain because they are given a chance to work out their ideas in writing, as well as in person-to-person discussions facilitated by the ideas generated originally in MeTAL.

The World Wide Web is a flexible tool that can easily be used in a classroom as tool for helping students write. However, this flexibility comes in the form of lack of structure for the content. There is no control over what hypertext links a user creates on a given page or how these pages or links are formatted. Instead, like most tools for creating traditional ITSs today, it is focused entirely on the media, the text and the graphics and the HTML, rather than on the writing. Students and teachers alike are easily distracted by having to write and sort through HTML documents. Final projects of classes often look pretty, with fancy fonts and images downloaded from far away places, but the writing is not improved from the reports of past years. This is mainly due to the fact that it is too easy to get distracted by surface features (making use of the felxibility inherent in HTML) rather than focusing on their ideas and their writing. It's the problem of form vs. content. Like the EON ITS authoring tool [Eon] and other knowledge-based authoring tools, MeTAL helps the users focus on the content of the domain where student learning is happening, rather than on the details of how to display that information.

MeTAL Hurdles - Computer Science Issues - Databases & Machine Learning

At the heart of the MeTAL system is not simple text files with HTML, but rather a complex object-oriented database. Because of this, it is possible to do many complex things that are helpful in tutoring. One of these, is doing queries on the objects, rather than simple full text searches as at most web sites. Simple HTML attempts to provide structure to documents, but it is not enforced, seldom used, and more often abused to simply get the right look. In MeTAL, we can easily queries such as: "show me all comments made by me since last week" or "show me all refutation nodes made that are linked to nodes with 10 or more votes" See [Idea Futures]. The system generates documents dynamically from a database and users may restructure their views according to their own preferences. For instance, nodes that have had many hits or votes or annotations associated with them could gravitate to the "top" or the "center" of the knowledge space. We think of the collaborative knowledge space as a graph or a true "web" rather than a hierarchical tree as is the case with Hyper-G systems [Hyper-G].

BEST COPY AVAILABLE

Using machine learning techniques, MeTAL will be able to generate "leading questions" for given nodes. This will be an aid in prompting users to By making use of the top indexed words from the entire collection of documents (as collected from model authors, or from past, fully implemented instantiations of a domain of interest) we can prompt the user for more information. The top indexed words from the collection can be compared against the top indexed words of the given author's documents/nodes. Words that are not mentioned can be used in the "leading questions" mentioned above. "Do you have anything to write about (your) _____ ?"

MeTAL Design Contributions

MeTAL is important because it an attempt to make the benefits of hypertext available to students in secondary school. The WWW and WWW browsers are great resources but are too unstructured and focused on media aspects rather than on the content -- the students's ideas and writing. It is not a replacement for person-to-person discussions in the classroom but is meant as a facilitating tool towards improved dialogues, especially in the large, 30 student classrooms we often have today. Without a way to bridge the gap between verbal communication and the writing of all students, many voices don't get heard. MeTAL is a way to facilitate this sort of communication. The WWW, and the internet in general, is said to be a decentralizing force and one that flattens the playing field, giving the little companies and individuals more opportunity for growth and power and voice. While this is true in isolated instances, in general, the WWW most benefits large corporations and business, rather than education and individuals. MeTAL is a step in the right direction by making use of the WWW for authentic writing and communication by students in their place-based community or classroom.

System References

[Annotations] Annotations. Group-annotations on the WWW, September 1993.

http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/group-annotations.html [Berners-Lee, et al. 1994]

[ASK] ASK http://www.ils.nwu.edu/info/software.html [Schank 1991]

[Belvedere] Belvedere http://info.pitt.edu/~suthers/ [Suthers 1995]

[CoVis] CoVis http://www.covis.nwu.edu/ [Gomez 1995]

[Cybernetica Principia] http://134.184.35.101/

[Eon] http://www.cs.umass.edu/~ckc/projects/eon.html [Murray & Woolf 1992]

[FEED] Feed Dialog http://www.feedmag.com/dialog.html

[Global Lab] Global Lab, TERC, Cambridge, MA http://www.hub.terc.edu/terc/gl/global-lab.html

[HYNECOS] HYNECOS http://www.cs.umass.edu/~ckc/691o/readings.html#8 [Vassileva 1994]

[Hyper-G] Hyper-G http://hyperg.tu-graz.ac.at/

[HyperNews] HyperNews http://union.ncsa.uiuc.edu/HyperNews/get/hypernews.html

[Idea Futures] the Foresight Exchange [Hanson 1990] http://www.ideosphere.com/

[LEGO LOGO] LEGO LOGO http://casr.www.media.mit.edu/groups/casr/papert.html [Papert 1993]

[Subury Valley School] Sudbury Valley School http://www.tiac.net/users/sdavid/svhd.html

[Threads] Hotwired Threads http://www.hotwired.com/threads/

[Xanadu] Xanadu http://www.aus.xanadu.com/

References

[Berners-Lee, et al. 1994] Berners-Lee, T., Caillau, R., Luotonen, A., Nielsen, H.F., & Secret, A. (1994) The World-Wide Web. Communications of the ACM, 37(8), 76-82.

[Gomez & Gordin 1995] Gomez, L. and D. Gordin (1995) A Case Study of Open-Ended Scientific Inquiry in a Technology-Supported Classroom; Proceedings AI-ED 95, 17-24.

[Hanson 1990] Hanson, R. (1990) Could Gambling Save Science? - Encouraging an Honest Consensus. Proceedings Eighth Intl. Conf. on Risk and Gambling, London, July 1990.

[Mander 1991] Mander, J. (1991) In the Absence of the Sacred: the Failure of Technology and the Survival of the Indian Nations, Sierra Club Books, San Francisco.

[McKendree 1995] McKendree, J. et.al. (1995) The Homeopathic Fallacy in Learning from Hypertext, ACM Interactions, July 1995, 74-82.

[McLuhan 1964] McLuhan, M. (1964) Understanding Media: the Extensions of Man, McGraw-Hill, New York.

[Murray & Woolf 1992] Murray, T. and Woolf, B. (1992). Results of Encoding Knowledge with Tutor Construction Tools, Proceedings 10th National Conference on Artificial Intelligence, AAAI-92, 17-23.

[Nelson 1995] Nelson, Ted. (1995) The Brown/MIT Vannevar Bush Symposium, October 12-13, 1995, ACM Interactions, March 1996, 61. http://www.ausbcomp.com/~bbott/wik/bushref.htm

[Papert 1995] Papert, S. (1993) The Children's Machine: Rethinking School in the Age of the Computer. New York: BasicBooks.

[Postman 1992] Postman, N. (1992) Technopoly: the Surrender of Culture to Technology, Knopff, New York.

[Schanck & Jona 1991] Schank, R.C. and Jona, M.Y. (1991) Empowering the Student: New Perspectives on the Design of Teaching Systems; Journal of Learning Sciences, 1(1), 7-35.

[Suthers, et al. 1995] Suthers, D. et al. (1995) Belvedere: Engaging Students in Critical Discussion of Science and Public Policy Issues; Proceedings AI-ED 95, 266-273.

[Talbott 1995] Talbott, S. (1995) The Future Does Not Compute: Transcending the Computers in Our Midst, O"Reilly & Associates, New York, .

[Vassileva 1994] Vassileva, J. (1994) A Practical Architecture for User Modeling in a Hypermedia-Based Information System; Proceedings 4th International Conference on User Modeling, August 1994, Hyannis, MA., 115-120.

[Wolf 1996] Wolf, Gary (1996) Steve Jobs: The Next Insanely Great Thing -- The Wired Interview, Wired 4.02, February 1996. http://www.hotwired.com/wired/4.02/features/jobs.html

[Wright 1995] Wright, Robert (1995) The Evolution of Despair, Time, August 28, 1995. http://pathfinder.com/time/magazine/domestic/1995/950828/950828.cover.html

Math and Science Gateway - A Successful Model for High School Educators and Students

Caroline Hecht, Cornell Theory Center, Cornell University, USA, hecht@tc.cornell.edu

Kathy Barbieri, Cornell Theory Center, Cornell University, USA, barbieri@tc.cornell.edu

July 29, 1996

Abstract: The World Wide Web provides new opportunities for education over the Internet. The Cornell Theory Center, a national center for high performance computing, has created an award-winning Math and Science Gateway for grades 9-12 (http://www.tc.cornell.edu/Edu/MathSciGateway/). It is designed to present the best resources from the Web, organized in a fashion familiar to both educators and students. The Gateway is updated frequently, with new materials being added, and outdated information removed. From this successful initiatve, other Gateways are being designed by high school teachers.

Introduction

The explosion of the Internet and the expanding number of Web sites provide a vast amount of information for educators and students. It can be very difficult for those unfamiliar with navigating the Net to locate organized information about specific topics. The Cornell Theory Center (CTC) has created an award-winning Math and Science Gateway for high school educators and students (http://www.tc.cornell.edu/Edu/MathSciGateway/), which provides annotated links to excellent math and science sites. These are organized by familiar topics, such as biology, chemistry, the environment, health, and mathematics. The Gateway also includes an extensive section for educators, and it provides an excellent entry point for those new to the Web and its wealth of information.

K-12 Education at CTC

For the past ten years, CTC has been a leader in developing and delivering high performance computing education. Because its audience is a national base of researchers, CTC has aggressively pursued online information delivery and most recently Web-based education. A variety of K-12 initiatives is included in our education program.

In 1989, CTC assumed the lead role in conducting the SuperQuest contest, a national high school competition designed to extend high performance computing to the high school level. Small student/teacher teams spent three weeks at a CTC summer institute, learning high performance computing techniques and working with Cornell faculty. Each winning school was given a workstation, which enabled the teams to continue their research. SuperQuest expanded in the following years and involved other partners and centers who offered similar summer institutes. A number of state and regional high school competitions have been developed from this successful model.

CTC sponsors Kids On Campus as part of our annual celebration of National Science and Technology Week. The purpose of this event is to increase computer awareness and scientific interest among local elementary school students. Children participate in numerous activities, including a hands-on session introducing them to the World Wide Web.

During 1995, we offered a series of workshops for K-12 educators and administrators. These workshops introduced them to the Internet and the Web. Included were talks describing resources for educators, and a hands-on lab in which participants explored the Web through the Math and Science Gateway. School groups may now request use of CTC's Training Facility to do training and hands-on exploration of the Internet and Web.

Description of the Gateway

In late 1993, CTC staff attending the Supercomputing '93 conference spoke to many K-12 educators who were interested in the Web but who found it unnecessarily confusing to locate the kinds of resources they needed for themselves and their students. In particular, it was clear that there was a gap on the Web in the areas of high school math and science, although plenty of relevant information was available for the younger grades. From this gap was born the idea of the Math and Science Gateway. Six months were spent collecting hundreds of potential resources and prototyping the Gateway. From there, CTC staff culled the best of those sites, organized and annotated them, separated out the resources that would be exciting to educators but not of interest to students, and produced the Gateway. It was announced in February of 1995.

The Math and Science Gateway provides an easy starting point for locating science and mathematics resources on the Web. It is tailored to the needs of students in grades 9 through 12, with links to resources in the subject areas of mathematics, computing, biology, chemistry, the earth, the ocean, the environment, meteorology, health, medicine, engineering, astronomy, and physics. The section for secondary school educators contains information on curriculum, software for the classroom, and Internet access in the schools.

The Gateway is optimized for ease of use by students. The resources accessible from the subject-area pages were selected to be particularly usable by high school students. This meant doing several things in the process of creating the Gateway. First, it was important to annotate every link. A page with links to "Welcome to the Planets," "Views of the Solar System," and "The Nine Planets" is of no use without descriptions of the sites to let you know which of the three you would find most interesting. Second, sites with specific information were preferable to sites containing long lists of links, only a few of which would be useful. For example, a site containing an article titled "About Temperature" was included, whereas a (hypothetical) meteorological organization whose home page included links to its bylaws, hiring policies, and a few dozen links to information about weather would not have been. Third, sites whose content or style were targeted for a much younger audience were omitted, even when those sites were very well done. Finally, sites with a commercial slant were omitted.

Eucators use the pages designed specifically for them in order to locate topics relevant to the classroom. For example, educators might find lesson plans, classroom activities, and curriculum databases, both for Internet-based and traditional education. There are links to discipline-specific resources, including articles, instructional activities, and inexpensive software for the classroom. Other resources for educators include information on how to contact educators at other schools to facilitate Internet-based collaborative projects for the students, how to get your school on line, summer programs, teachers' organizations, and electronic discussion lists. In addition, educators peruse the subject-area pages targeted for students, to explore the multitude of resources included there.

Awards, Feedback, and Usage

The Gateway has received many compliments and comments, as well as awards and write-ups.

It was rated among the top 5% of all Internet sites by Point Survey in the Education / K-12 category, and has received a 3-star award from the McKinley group. It won the award of "Excellence" in the 1995-96 WWW Home Page Competition sponsored by the Lone Star Chapter of the Society for Technical Communication. It was listed in the "Best of comp.infosystems.www.announce" when it was announced in that newsgroup.

The Gateway has been described in articles in "Mathematics Teacher" (Doerr and Hecht, 1995) and in "American Scientist" (Mike May, 1995). It was included in lists of Web sites for educators in the London Times Educational Supplement and in the May/June 1996 "National Council of Teachers of Mathematics News Bulletin."

We have received much feedback indicating that the Gateway serves its purpose and fills its niche on the Web very well. In addition to numerous non-specific comments (e.g. "Love this site, thanks for creating it."), we've heard from students and educators using the Gateway in a variety of ways.

From February through June 1996, we included a questionnaire on the Gateway's home page to solicit more specific feedback. We received approximately 400 responses, about half from students and half from educators. Among our findings was that many Gateway users are relatively new to the Web, affirming our original feeling that new users needed a Gateway such as this to help them navigate the multitude of resources that exist on the Web.

We received dozens of questionnaires from 7th grade students in Tully, NY, prompting us to get in touch with their teacher to find out how they were using the Gateway. The teacher had incorporated the Gateway into the science curriculum, and the students were using class time to explore resources in biology, health, chemistry, and the environment.

A math teacher from Ohio wrote, "I intend to recommend this to my Advanced Math/AP Calculus students in working on class reports and investigations, and for the researching of an annual term paper (topics such as the Golden Ratio, Fibonacci Sequence, Fractals, and Cryptology)."

Other teachers are using the Gateway to find resources to enhance the curriculum for basic and advanced science classes, to improve teachers' abilities in the classroom, and as a way to introduce both educators and students to the Web. Students have told us that it's useful in their research and that it helps with homework assignments; here are a few of their comments:

- "I use the Gateway to find out more about different mathematicians." (10th grade)
- "I'm in a Co-op researching information for teachers to use for their classes. These links have been a great help to me. Thanks." (12th grade)
- [How the student uses the Gateway] "To study for an upcoming test." (9th grade)
- "This helps me find what I want in sciences and helps me improve my math and science skills." (10th grade)

The Gateway has averaged close to 15,000 hits per month. Most of the subject-area pages get 200-400 hits per month, except for the math page, which averages about 2000. We have attempted to determine why this one page is so much more popular than the others, and have concluded that there are more collections of high school math resources on the Web than for other subject areas. Many of these collections link to the Gateway or directly to its math page. We can only surmise that mathematics educators have unusually high interest in using the Web.

Maintenance of the Gateway

A Gateway such as this would quickly become obsolete if it weren't being updated regularly. CTC staff are on the constant lookout for new resources to add. In addition to perusing standard Web announcement lists, such as the newsgroup comp.infosystems.www.announce, Gleason Sackman's net-happenings listserv, and the weekly Scout Report, CTC solicits the help of local educators who keep us posted when they find relevant new Web resources.

We also receive unsolicited mail from our users, with their suggestions for additional resources. These are meticulously checked by CTC staff for their suitability. Roughly 80% of the suggestions to date have been added. The other 20% are mainly sites that are designed for a different age group (elementary school students, college students), sites that are too commercial, and sites that have too little educational content.

Knowing how frustrating it is to find broken links, CTC staff manually check each and every link in the Gateway every few months. The advantage of having a person doing this rather than a program is that we can catch URLs whose content has changed and are no longer appropriate, as well as those which have moved and left a "forwarding note" behind.

With the Gateway less than two years old, it hasn't yet grown too large. However, with the rapidly increasing number of Web resources, the Gateway has the potential of expanding to the point where it will be difficult to navigate. So far, rather than deleting some resources when we add new ones, we've

BEST COPY AVAILABLE

made an effort to keep it organized by creating new subcategories when necessary.

From One Gateway, Another is Born

CTC has used the Gateway as a home page for educators and students attending Internet workshops. During the first of these workshops, in March of 1995, two local school librarians approached us and expressed interest in creating a companion Gateway in Arts and Social Sciences. With CTC's assistance, they completed that project later in the year. The two Gateways now include links to each other, and the pages for educators have been expanded to include additional resources of interest to educators outside of math and science.

Summary

A Gateway optimized for ease of use is essential for those being introduced to the wealth of information available on the Web. Organizing the Gateway by familiar subject areas proved to be user friendly. Maintenance of the Gateway is also critical, because new Web resources appear daily. This approach has been very successful, and has provided a model for the creation of gateways in other subject areas.

Summary of URLs

CTC Math and Science Gateway - http://www.tc.cornell.edu/Edu/MathSciGateway/
CTC Arts and Social Sciences Gateway - http://www.tc.cornell.edu/Edu/ArtSocGateway/
CTC Kids on Campus Web Demonstration - http://www.tc.cornell.edu/Kids.on.Campus/WWWDemo/
CTC Education Programs - http://www.tc.cornell.edu/Edu/
CTC Home Page - http://www.tc.cornell.edu/

Welcome to the Planets - http://pds.jpl.nasa.gov/planets/
Views of the Solar System - http://bang.lanl.gov/solarsys/
The Nine Planets - http://seds.lpl.arizona.edu/nineplanets/nineplanets/
About Temperature - http://www.unidata.ucar.edu/staff/blynds/tmp.html

Net Scout Services (The Scout Report and Net-Happenings) - http://rs.internic.net/scout/
Point Survey - http://www.pointcom.com/
The McKinley Group - http://www.mckinley.com/
STC, Lone Star Chapter - http://stc.org/region5/lsc/www/ipic/wwwenter.html
Best of comp.infosystems.www.announce - http://www.boutell.com/announce/index.html

References

[Doerr and Hecht, 1995]. Doerr, Helen, and Hecht, Caroline (November, 1995). "Technology Tips: Navigating the Web," *Mathematics Teacher* 88, no. 8, pages 716-719.

[Mike May, 1995]. May, Mike (November-December, 1995). "Scientific Surfing for Children," *American Scientist* 83, pages 568-571. This is also available at URL http://www.amsci.org/amsci/articles/articles-MMay.html

Acknowledgments

Kathy Barbieri developed the concept of the Gateway based on her experiences at Supercomputing '93 and worked with Tony Gonzalez-Walker to develop the initial prototype. Caroline Hecht created the current version of the Gateway and continues to maintain it. Helen Doerr provided guidance and support throughout.

The Cornell Theory Center Arts & Social Sciences Gateway was conceived of and created by two school librarians in the Ithaca area: Jo-Ann Mancini, librarian at Lansing High School, and Barbara Nosanchuk, librarian at Cayuga Heights Elementary School; they received help from Caroline Hecht, in the CTC Online Information Group.

MITRE Information Discovery System

Daniel J. Helm, Raymond J. D'Amore, Puck-Fai Yan
Digital Libraries Group
The MITRE Corporation
1820 Dolley Madison Blvd., McLean, VA 22102-3481, USA
{dhelm, rdamore, pyan}@mitre.org

Abstract: The MITRE Information Discovery System (MIDS) is a baseline system for integrating advanced processing tools for information discovery and retrieval in large-scale distributed environments. The system is built on a modular, extendible architecture that allows for system-level decoupling and allocation of component processing tools across network nodes to provide for efficient processing in distributed environments. At one level, the system provides for multi-platform user access to HTTP, Gopher, FTP, and news servers using an HTML based client interface. However, more significantly, the system provides advanced tools for metadata generation from disparate network objects, and a content routing mediation layer for classification of metadata into appropriate information brokers. This bottom-up layered information organization approach supports a wide range of information retrieval and browsing strategies.

Introduction

The MITRE Information Discovery System (MIDS) project is a multi-year MITRE-sponsored research program to develop a set of multi-faceted capabilities for collecting, categorizing, organizing, and discovering digital information in a distributed environment. MIDS tools not only support information management and retrieval, but post-retrieval operations for summarizing and presenting information to the user.

A white pages directory organizes items by name (e.g., Internet user names), while a yellow pages organizes information by attributes. MIDS is intended to support a dynamic yellow pages capability; that is, information is organized by "topic," and can be dynamically reorganized as information collections change or new collections are added. The system consists of an information organizer that groups multiple collection contents topically, and a multi-broker network that uses object summaries to create a type of adaptable subject catalog that is used to support browsing and searching specific collections. The organizer and broker are the primary components which provide the foundation for the MIDS architectural framework.

The main emphasis of the program so far has been on the effective integration of advanced information retrieval and natural language processing capabilities since there are a number of important system integration issues and effectiveness and efficiency concerns that are critical to developing even a rudimentary pilot system. While the short-term objective is not to support a large-scale system evaluation such as currently being supported by ARPA's Tipster program, the current work will be focused on selected operational-based, end-user assessments of system effectiveness and efficiency.

MITRE Information Discovery System

Essentially, MIDS is organized around several key functions: collection, indexing, routing, organization, storage, retrieval and browsing. These functions are typically handled as tightly integrated components, in traditional commercial off-the-shelf systems; however, MIDS provides these functions as separate modules that can communicate within a networked environment. This distributed, modular approach provides flexibility in terms of the methods employed and in defining extensions to the baseline architecture.

The MIDS project leverages heavily of both MITRE-developed technologies and tools available in the public domain. As a result the system incorporates technologies for metadata extraction, clustering, filtering, retrieval, and document summarization with new development focused on specialized problems not readily addressed by off-the-shelf capabilities. In a sense, MIDS is both an integration framework and a set of tools. As an example, the integration framework is centered around the Harvest [Bowman et al. 94] system.

Harvest is a scalable, customizable discovery and access system developed by the Internet Research Task Force Research Group on Resource Discovery (IRTF-RD). Major Harvest components utilized include Gatherers and Brokers. The Gatherer is a subsystem which collects metadata (indexing information) from a set of providers (file servers). The Gatherer knows about different object types (e.g., HTML, text, PostScript) and is able to parse and extract different metadata depending on the type of object being processed. A Broker provides a search interface to information collected by one or more Gatherers or other Brokers. Harvest provides much of the system-framework for handling distributed information. However, as discussed below, the basic Harvest system does not address well requirements for content routing to multiple Brokers where information cannot be a priori linked to a particular Broker, nor can it handle effectively query routing where broker level metadata is needed to effectively assign user queries to appropriate Brokers. These areas and others motivated a number of extensions to the basic Harvest system.

System Architecture

The MIDS system architecture is given in [Fig. 1]. The Gatherers are Harvest subsystems which extract metadata from one or more providers (sources of information). The metadata format used, called the Summary Object Interchange Format (SOIF) is based on a combination of the Internet Anonymous FTP Archives (IAFA) IETF Working Group templates [Deutsch et al. 94] and BibTex [Lamport 86]. It has an attribute-value format which is easily parsed and yet sufficiently expressive to handle many kinds of objects. The Gatherer Dissemination Service (GDS) periodically collects streams of metadata in SOIF format from a set of Harvest Gatherers. Using the collected metadata, the GDS provides a content routing capability that classifies the documents associated with the metadata into a set of topical areas, and builds a set of files containing streams of SOIF records for Brokers that had registered themselves with the GDS. It also disseminates its processed information to Brokers and the Broker Information Service (BIS). Brokers are Harvest subsystems that provide a full-text search capability for MIDS. Each Broker manages a set of topical areas and is periodically updated with SOIF records from the GDS based on the information type it desires. The BIS provides a topical browse service, enabling a user to retrieve documents based on topic selection. It also provides a query routing [Sheldon et al. 94] capability by determining which Brokers to search for a particular topical area. The client interface provides a user interface to enable effective navigation, browsing, and searching of information.

Information Collection

Information is collected into MIDS by utilizing Harvest Gatherers. The GDS contains a list of Gatherers that it periodically contacts for streams of metadata. This information is stored in an input queue on disk for later processing by the classification engine which is part of the GDS. For better network utilization (bandwidth), it is more efficient to run Gatherers at each provider site, although Gatherers can access data remotely as well. Running a Gatherer locally enables it to collect metadata by directly accessing documents via file system I/O, and then have it ship one compressed file of its processed information to the GDS. Running a Gatherer remotely requires that it obtain each document through the HTTP, Gopher, or FTP protocol, which will incur a much greater performance penalty.

Information Classification

There are two methods used currently to classify information within the GDS. To provide for a tailorable classification scheme, tools are provided for a knowledge engineer to define a knowledge base of classification rules based on a fixed taxonomy of topics. The knowledge engineer will define the relationship between topics of interest as well as descriptors used to map documents into the topical hierarchy. The classification engine used to process information into the fixed taxonomy is a SIFT [Yan et al. 95] filtering engine, modified to provide for a more expressive query language, and also to process streams of SOIF records as opposed to regular full-text documents. SIFT's output processing logic was also modified to generate more classification results, and also to generate Broker specific output files containing streams of SOIF records matching Broker interest profiles.

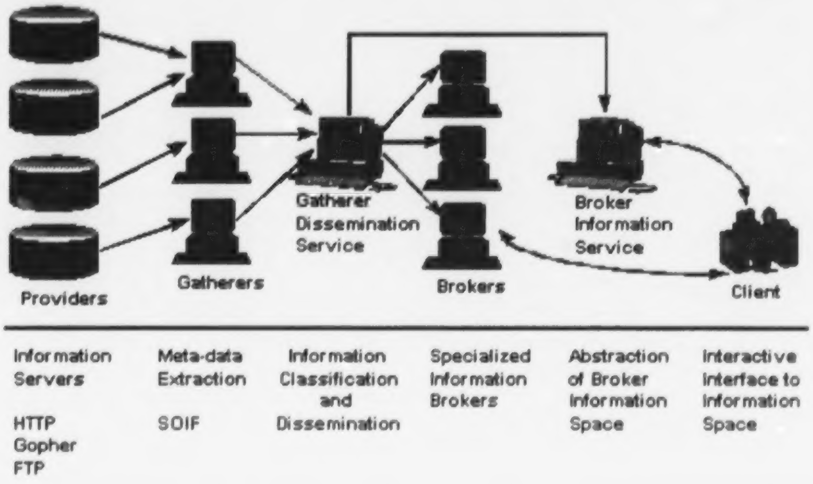

Providers
Gatherers
Gatherer Dissemination Service
Brokers
Broker Information Service
Client
Information Servers
HTTP
Gopher
FTP
Meta-data Extraction
SOIF
Information Classification and Dissemination
Specialized Information Brokers
Abstraction of Broker Information Space
Interactive Interface to Information Space

Figure 1: MITRE Information Discovery System (MIDS) Architecture

To augment the fixed taxonomy classification scheme, several clustering schemes (e.g., [Jain et al. 88]) are under investigation to provide a topical decomposition of specific subject categories. After information gets filtered into the fixed taxonomy, it is processed through a clustering algorithm which groups documents in each topical area into related piles. This is done to provide a finer granularity of topics than is supplied by the fixed taxonomy classification scheme. This is especially useful when a large number of documents are assigned to a topical area as a result of the filtering process.

Information Organization

In addition to collecting and classifying information, the GDS distributes its processed information to a set of Harvest Brokers and the BIS. Brokers register themselves with the GDS and have profiles, stored in a knowledge base, that specify the topical categories of the kinds of documents to be received. Each time the GDS classifies new information, Broker specific files containing SOIF records are written to an output queue on secondary storage. Brokers periodically contact the GDS to receive new information and to index it. In addition, information pertaining to objects and their classification categories is routed to the BIS.

Changes that occur to the document collections processed by the GDS, in terms of document deletions, additions, and updates, are reflected throughout MIDS using a weak-consistency protocol [Downing et al. 90]. Since each MIDS subsystem is run only periodically (weekly), changes occurring to document collections are not reflected until a new batch of information is processed through the system. After processing a new batch of information, the GDS notifies Brokers and the BIS about any changes that had occurred so that they can appropriately update their databases.

Information Discovery and Retrieval

Information discovery and retrieval in MIDS occurs through the client interface which interacts with Harvest Brokers and BIS. The initial services provided by the system include topical browse, query routing, and search. The BIS provides a topical browse service by managing the fixed taxonomy of topics as well as the topics generated as a result of utilizing clustering techniques. The BIS also manages document summary information, which had been generated as a result of the classification process in the GDS, and provides the query routing capability by determining which Harvest Brokers to search as the user browses the topical information space. This is accomplished by means of a table which lists the association between Brokers and the topical categories they manage.

Harvest Brokers provide a full-text search capability, enabling a user to issue a query containing a Boolean expression of keywords and then receive a list of documents.

All documents retrieved in the system reside at the information provider sites. Only metadata is processed and stored within MIDS. After the user consults the system, a list of documents represented by titles and URLs (Uniform Resource Locators) are presented. The user can then select a document, which subsequently gets retrieved from the appropriate server (e.g., HTTP, Gopher, FTP) at the provider site.

Post-retrieval Tools

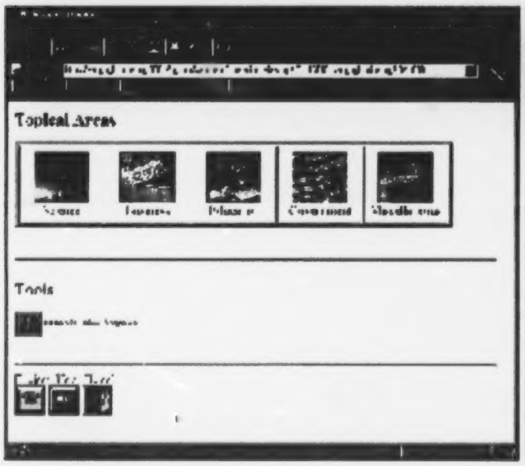
The system is designed to support a wide range of post-retrieval tools to help users process information after it is retrieved. The main services provided include document summarization and abstracting. These services utilize a Part-of-Speech (POS) tagger based on the work of Eric Brill [Brill 93], which has been modified by MITRE for better performance. Processing documents with a POS-tagger allows us to abstract and summarize based on parts of speech. In the initial capability, a tagged document can be filtered for verb and noun forms thereby reducing the word volume for the document. We eliminate common verb forms to avoid biasing results. The remaining verb and noun forms can then be totaled to produce a vector for the document. This vector can be used for clustering or passed forward to the summarization process. The summarization process examines the verb and noun vector for the document and selects the most used words in the document. Those words are then used to score each sentence of the document. The highest scoring sentences are then presented as a summary in the order in which they appeared in the original document. The number of sentences chosen is relative to the size of the document, but will not exceed a maximum threshold.

Experimental System

We have built a prototype MIDS to demonstrate our ideas and to evaluate and extend the system. The system currently collects and processes approximately 20,000 documents from a number of servers at MITRE. We currently have one Harvest Gatherer collecting information from three MITRE servers, four Harvest Brokers providing the full-text search functionality, one GDS, and one BIS. Within the GDS classification knowledge base, we have defined a number of fixed taxonomies covering the areas of business, government, science, and education. Within the area of science, we have incorporated a subset of the ACM Computing Reviews Classification System.[1]

Example Session

Our prototype system initially presents the top level screen shown in [Fig. 2] to the user. The screen displays the high level topical categories as well as a list of tools including "Search" (full-text search). [Fig. 3] shows a subset of the topics displayed as the result of selecting the "Science" topical area. The list box presents the hierarchy of topics beginning with the selected root category. Indentation is used to show a topic, sub-topic relationship. The number to the right of each leaf topic is the number of documents assigned to that area. A user can select one or more topics from the list box, select the "View Documents" button, and receive a breakdown of all documents in each area selected as shown in [Fig. 4]. The score associated with each document in the list represents how similar the document is to the topical profile. Optionally, a user can specify "List groups" under "Results Type" to receive a clustered listing of documents associated with selected topics. [Fig. 5] shows a sample clustered view of documents. Each cluster of documents is preceded by a list of the top discriminating terms, with the left-most terms being the highest discriminators. The score associated with each document in a cluster, corresponds to how similar the document is to the discriminating term list (cluster centroid). The user can then click on a document title to have it retrieved and displayed from the provider site, with topical descriptors highlighted in the text (Fig. 6); or click on "Summary Information" to see a summary of the document. The summary includes a list of other topics assigned to the document as a result of the classification process, and an abstract of the document generated by processing it through the Brill tagger software. [Fig. 7] shows a sample document summary. Within a topic list screen, a user can select the "Search" function to perform a keyword search against all Harvest Brokers that manage information associated with the selected topical area. The query routing service provided by the BIS determines the appropriate Brokers to query, where results from different Brokers are combined into a single list similar to that shown in [Fig. 4].



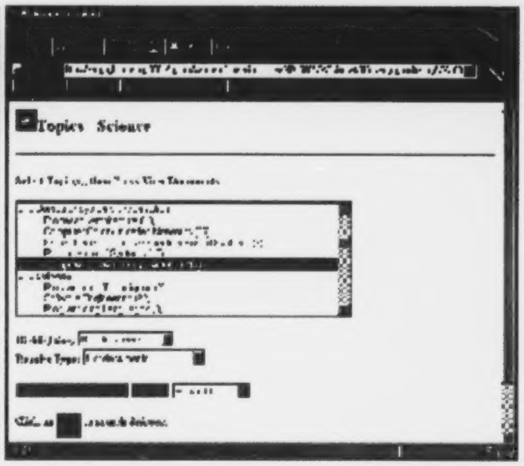
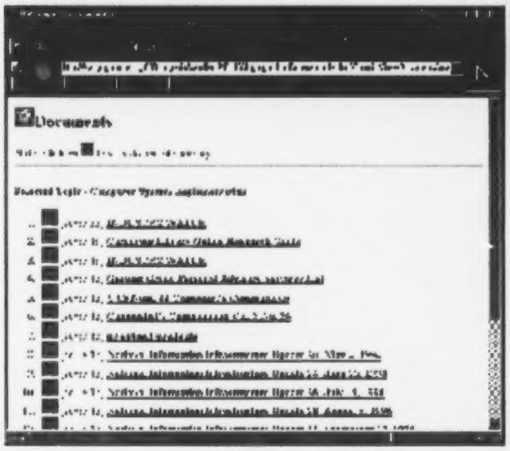


Figure 2: Top Level Screen **Figure 3:** Topic List Screen



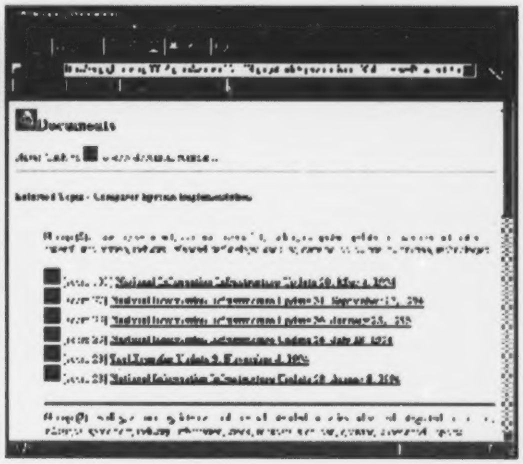
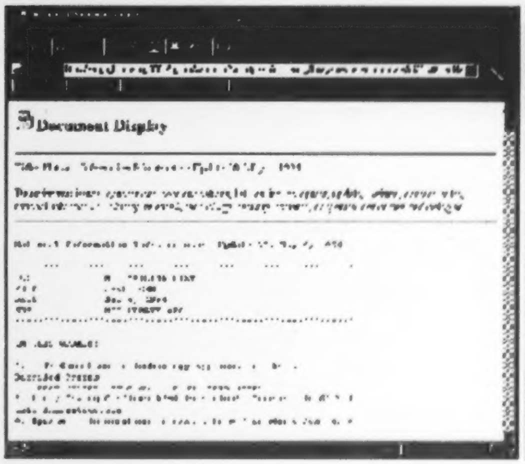


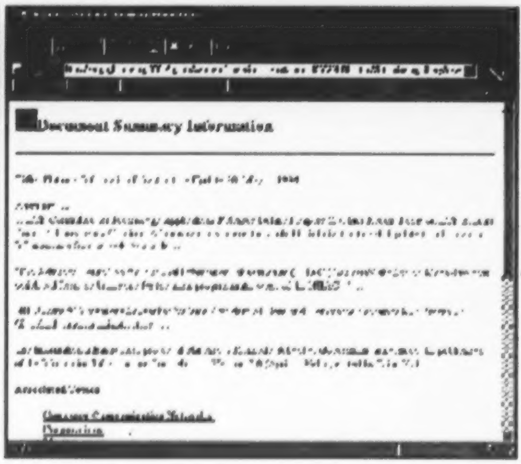
Figure 4: Document List Screen **Figure 5:** Clustered List Screen

Conclusion and Future Work

This paper describes ongoing work to build a resource discovery system that provides services for collecting, categorizing, organizing, and discovering information stored in a distributed environment. Beyond its potential use in operational environments it has further value as an extendible testbed for incorporating additional advanced information processing technologies as they become available. As such, the system provides a framework for evaluating both enabling information handling technologies and a basis for assessing integration strategies. More specifically, planned enhancements include extending the Harvest Gatherer to provide a range of new services to include generation of a '"richer" abstraction of an object's content, to provide built-in methods for handling access controls, security, and cost accounting. These extensions will provide a "Smart" Gatherer capability that will be more viable in disparate information environments; especially where information is managed in autonomous collections and cannot be "freely" accessed. Scaling issues will provide another major focus and will include efforts directed at making the system more scalable through the use of replicated components (e.g., GDS and BIS). Finally, work will focus on developing the post-retrieval and pre-retrieval toolsets that will be used to provide query enhancement capabilities, visualization tools, and personal information management capabilities.



Document Display



Document Summary Information

Figure 6: Document Display Screen **Figure 7:** Document Summary Screen

References

[Bowman et al. 94] C.M. Bowman, P.B. Danzig, D.R. Hardy, U. Manber, and M.F. Schwartz. Harvest: A scalable, customizable discovery and access system. Technical Report CU-CS-732-94, Department of Computer Science, University of Colorado, Boulder, Colorado, August 1994.

[Deutch et al. 94] P. Deutsch, and A. Emtage. Publishing Information on the Internet with Anonymous FTP. Bunyip Information Systems Inc., May 1994.

[Lamport 86] L. Lamport. LaTeX: A document Preparation System. Addison Wesley, Reading, Massachusetts, 1986.

[Sheldon et al. 94] M.A. Sheldon, A. Duda, R. Weiss, J.W. O'Toole, Jr., and D.K. Gifford. A content routing system for distributed information servers. Proceedings of Fourth International Conference on Extending Database Technology, March 1994.

[Yan et al. 95] T.W. Yan, H. Garcia-Molina. SIFT - A Tool for Wide-Area Information Dissemination. Proceedings of 1995 USENIX Technical Conference, January 1995.

[Jain et al. 88] A.K. Jain and R.C. Dubes. 1988. Algorithms for Clustering Data. Prentice Hall.

[Downing et al. 90] A.R. Downing, Ira B. Greenberg, and Jon M. Peha. OSCAR: an Architecture for Weak-consistency Replication. Proceedings of IEEE PARBASE-90, March 1990.

[Brill 93] E. Brill. A Corpus-Based Approach to Language Learning. Ph.D. thesis, Department of Computer and Information Science, University of Pennsylvania, 1993.

BEST COPY AVAILABLE

The Impact on Education of the World Wide Web

Dr DJ Hobbs
School of Computing, Leeds Metropolitan University, UK
d.hobbs@lmu.ac.uk

RJ Taylor
Distributed Systems Technology Centre, Queensland University of Technology, Australia
taylor@dstc.qut.edu.au

Abstract: Although interest in the WWW as a way to teach and to empower students is increasing, on-line classrooms are still few and far between. The newer Internet technologies may help change this. For example, developers are now experimenting with technologies which can provide users with synchronous communication, using video cameras to allow easy access to live discussions in a WWW environment.

This paper describes a project which created a set of WWW pages documenting the state of the art in educational multimedia design, and then went on to design, develop and evaluate a prototype WWW-based multimedia teaching tool - a Podiatry test using HTML forms, 24-bit colour images and MPEG video. The project was conducted between sites in the UK and Australia using the WWW and other Internet functionality to exchange and share information, giving valuable first-hand experience of the benefits and frustrations of on-line cooperative working.

New Internet Functionality

The WWW has found enormous success through word of mouth coupled with the enthusiastic support of the media. Probably more importantly, commercial developers have now adopted the WWW as a new advertising medium. Along with this increase in popularity, there has been a rush to incorporate new features into the software, many of which can facilitate and enhance specific educational modes discussed later in this paper. For example, as HTML has evolved, new features have been added such as interactive forms, defined "hot spots" in images, more versatile layout styles, and formatted tables. An interesting area is the integration of new Internet technologies into the WWW. For instance, Internet Relay Chats (real-time group discussions) and MUD's (Multiple User Dungeons, essentially an IRC in an interesting setting) provide users and designers with the ability to interact with each other live instead of, for example, having to wait for a mailing list to distribute the information as it is posted via e-mail. Another Internet technology being integrated into the WWW is CUSeeMe sites (ftp://gated.cornell.edu/pub/video). CUSeeMe is a teleconferencing program which allows users to see and hear each other by converting the data from a video camera into an Internet compatible format. Functionality such as this combined with the WWW's built-in audio-visual capabilities suggests new possibilities for group-based on-line education.

Another influence likely to affect the WWW in a major way is HotJava, a WWW browser developed by Sun Microsystems (http://java..sun..com/) that can execute programs (known as applets) written in the Java programming language and included (like images) in HTML pages. The Java language is the first to present a comprehensive solution to the challenges of programming for the Internet, providing an object-oriented programming language optimised for the creation of distributed, executable applications and offering portability, security, advanced networking and reliability without compromising performance. Because Java is compiled into machine independent bytecodes, applications written in Java can migrate transparently over the Internet, accessible by anyone using the HotJava browser or any other Java enabled browser. Netscape Navigator, by far the most popular WWW browser, now offers Java support enabling it to download Java applets to run on a wide variety of client systems. Java has been available since April 1995 for developers using the Sun Solaris or SunOS platforms. More recently the availability of the language has been extended to other varieties of UNIX, Microsoft's Windows NT and Windows 95 operating systems.

The RealAudio (http://www.realaudio.com/) player is a another good example of how the WWW technology is being advanced. It gives WWW users equipped with only a standard multimedia personal computer and a 14.4 Kbps modem instantaneous access to real-time audio programming. It uses highly compressed audio files which, rather than being downloaded and then played as a complete file, are received incrementally with each segment fed to the RealAudio player as it arrives thus enabling continuous play from the outset. This technology makes audio-on-demand as practical as delivery of text and graphics, overcoming the significant downloading delays that have hitherto presented an obstacle to its informational, recreational and creative use.

The latest extension of the WWW browser is the development of "plug-ins". These application are written by third parties and extend the functionality of the browser offering features such as video and audio compression and decoding, virtual reality markup language (VRML) support and video/audio conferencing within the browser's Graphical User Interface.

The WWW for Research and Information Seeking

The WWW has grown from its origins as a simple system for distributing documents and communicating among members of the high-energy physics community into a more general multimedia tool of wider appeal. Nevertheless, its use as a research tool continues and has increased enormously. Research institutions and universities have established home pages and the WWW is increasingly used to advertise the work and interests of departments and staff. Researchers are still some of the keenest users of the WWW in view of its potential for contact and collaboration, for disseminating research findings, and for facilitating peer review of the outcomes of research [Russell & Baird 1995].

Despite the magnitude of the information available, using the Web is a simple matter, even for young children, so that anyone with the desire and the time to explore is bound to discover a variety of fascinating sites and resources. Yet because of the vast numbers of WWW sites, knowing where to begin searching, what to look for, and what to ignore can be a daunting task. The development of what are known as webworms, spiders, and knowbots (computerised search agents which will scan the Internet looking for requested information) has facilitated searches somewhat although it has brought another problem in that an overwhelming amount of reference material may be returned. This has been helped somewhat by individual WWW users who have already begun to catalogue the enormous variety of educational resources available on-line. As this information is compiled, it becomes available to the educational community in the form of on-line resource guides, some of which are no more than hypertext lists of known educational resources.

The educational resources page designed and constructed as part of this project is an example of this kind of listing. A set of WWW pages (http://www.dstc.edu.au/AU/staff/richard-taylor/education/) was developed to offer the reader a practical means of following up the educational resources discussed. The pages were designed and constructed featuring sites demonstrating interesting and innovative educational pages and existing on-line courses which were considered to exploit some of the best educational features of the WWW. As designers of WWW sites develop more on-line resource guides, the WWW will begin to resemble a World Wide Library Catalogue although, unlike a traditional library, the books and other documents that are available will have been created by students, lecturers and anyone else with an interest in contributing. In consequence, as the WWW grows, so will easy access to useful and interesting information. There are naturally concerns over the reliability of the knowledge that is being published on the WWW. [Treloar 1995] suggests, however, that there is no reason why Web-published journals should not be subject to the same peer review processes that apply in the print world.

The WWW as Teaching Tool

From a curricular point of view, the WWW can be used to design tutorials and on-line lessons for a variety of subjects. For example, [Blumberg 1994] describes an on-line teaching tool for basic genetics known as MendelWeb (http://netspace.org/MendelWeb/) that integrates elementary biology, discrete mathematics, and the history of science. MendelWeb is constructed from Gregor Mendel's 1865 paper *Experiments in Plant Hybridisation* presented as an active document, with links to traditional reference material (e.g. glossaries, biographies, and the original German text) as well as images, tutorials, active commentaries, related Web sites, and animations. Discussion and questions are presented as they would be in a live introduction to biology course - students may choose in what order they wish to explore the topics, enabling them to develop their comprehension of the subject at their own rate.

The authors of such systems have usually carefully mapped out the possible outcomes of each piece of information offered. A successful WWW system such as MendelWeb will therefore be crafted with sometimes thousands of links and hundreds of pages. The proliferation of automatic HTML authoring programs suggests that creating such linkages in the future will no longer seem a daunting task. Furthermore, hyperbook design may become even further simplified as a result of work by programmers at Brown University who are working on what are known as ASK systems - automatic, intelligent computer programs which will analyse a document's content with inquisitive search agents in order to help formulate questions that might be raised by that content. The WWW is therefore beginning to provide the necessary tools with which to design on-line teaching material. However the potential of WWW teaching packages has yet to be realised largely because most WWW books have been technically oriented. In order for this technology to reach the mainstream subject areas, WWW tutorials must be designed for less technical subjects like history, music, language, arts. For example, a site specialising in the complete works of Shakespeare site could include question and answer sessions, as well as audio and video clips of each play and poem. Using appropriate tools, the capability to transform a topic of choice into a document that could be useful and educational for students could be available to everyone.

The WWW as Examiner

The developments in HTML that have introduced the ability of the WWW to display fill-in forms can also help create interactive educational pages. Educational sites have begun offering tests and quizzes for both assessment and self-assessment. The Podiatry test developed during this project and described in more detail below is an example of this kind of application of the WWW. However, it is not yet generally productive to produce this kind of interactive software on the WWW. For example, in the area of medicine which is very visually intense and information rich [Cho 1994], it is not easy for specialists with little training in computer science to write scripts that create and manage forms and interact with other software on the server. Another technical problem discussed by [Ibrahim & Franklin 1995] in relation to these kinds of interactive applications is the fact that the http protocol is stateless in the sense that there is no direct relationship between two consecutive requests to the same server, even if the queries come from the same user. The server treats every request it receives independently from any other request it received in the past or that it will receive in the future. From a technical perspective, this statelessness allows the http server software to impose very little overhead on the server machine, and keeps the protocol between the client and server very simple. Nevertheless, from a learning point of view the statelessness of the http protocol (meaning that the connection to a server is closed after a requested document is delivered) is a serious shortcoming preventing intelligent interaction.

The WWW as an Educational Forum

Proponents of the Internet have long promoted its use as a forum for discussion and as a marketplace for ideas and information. The WWW also fulfils this goal, and in terms of use in the education community, the WWW can provide a basis for virtual debate and discovery. All of the original uses of the Internet - including file transfer protocol (ftp), e-mail, USENET news, and gopher continue to thrive in the context of the WWW and have now converged into a singular informational tool, since the latest generation of WWW browsers, such as Netscape Navigator are capable of interacting with the full suite of Internet protocols. Because of this, it is conceivable for a designer to utilise all of these services to set up a multimedia/hypermedia discussion on any given subject. As a basis for such discussions, Internet users have traditionally used mailing lists to form a discussion group, receiving information from and posting information to the mailing list via e-mail which in turn then distributes the information to everyone on the list. Recently, some organisations have even used mailing lists to run virtual conferences, where sometimes thousands of people sign up to an on-line discussion and take part in a week-long forum, all without leaving their homes or offices.

For educators, this combination of presentation (the WWW) and critique (mailing lists) can be used successfully in a variety of ways. For instance, a teacher could set up a WWW site which comprises the lectures, frequently asked questions, and multimedia presentations. Via a mailing list, students could automatically add information to that site in the form of additional questions, reports, essays, etc. In this case, they could use mail-to-HTML converters and so would not need to become experts in HTML. The software would then automatically append their message onto the page itself, so future site visitors will be able to read the comments. The educational potential of such a system cannot be ignored although the uncontrolled and unrefereed nature of the material again has to be born in mind.

Students can also use the various Internet technologies to create their own hypertext work and then present it on-line so that their peers and lecturers may discuss and review it. Learning how to critique others' work and to present a persuasive, constructive argument are skills that are often gained slowly for many students, for they are rarely taught in any formal fashion [Laurillard 1995]. Further, on-line electronic discussions apparently seem to be less threatening for some than standing up in front of peers. In addition, because conversation is electronic, it can be automatically catalogued and presented by the student as part of the project. This is not to propose that traditional class presentations should vanish with the advent of on-line class forums, but that allowing students to work with and learn from each other in such a way could encourage the many students who previously did not easily contribute voluntarily to a discussion.

The WWW in Collaborative Education

In its present form the WWW is not very well suited for collaborative work which requires a high degree of real-time interaction. Prototypes for synchronous communication such as Web-Chat exist but are currently unstable and slow. One of the most exciting recent developments for collaborative education are WWW interfaces to Multi-User Dungeons (MUD's) and Object-oriented MUD's (MOO's). [Curtis & Nichols 1993] describe a MOO as a network-accessible, multi-user, programmable, interactive system well-suited to the construction of text-based adventure games, conferencing systems, and other collaborative software. Its most common use, however, is as a multi-participant, low-bandwidth virtual reality. Participants give one-line commands that are parsed and interpreted as appropriate. Such commands may cause changes in the virtual reality, such as the location of a character, or may simply report on the current state of that reality, such as the appearance of some object. The database contains representations of all of the objects in the virtual reality, including the MOO programs that the server executes to give those objects their specific behaviours.

MOO's allow for individual users to extend the environment by "building" or creating new objects. In an educational context this can allow the student to become an active participant in the learning experience. In addition, it is well documented in the literature that MOO's provide a strong sense of "place", possibly bringing back some of the social intercourse of "campus" life that is lost in distance education. A MOO server can also be configured to act as an http server. This means that a WWW browser can be used to look at locations, rooms, people, artefacts, etc. in the MOO. These objects can have hypertext URL's attached and therefore be used to structure information on the WWW. An example use of a MOO in an advanced educational setting is the Global Network Academy (GNA) 'Introduction to Object Orientated Programming using C++' course (http://info.desy.de/pub/uu-gna/html/cc/index.html). The long term goal of the GNA is to become a fully accredited on-line university. The course was built around four main components, a hyperbook, mailing lists, practical projects and MOO interactions [Perron 1994].

A Demonstration of the Educational Potential of the WWW - The Multimedia Based Teaching and Assessment Tool

The process of exploring the WWW is in itself a educational experience. There are however more structured ways that the WWW can be used in the educational context. With the continuing development of HTML and its ability to use display fill-in forms, courseware designers are now able to create educational material that has most of the characteristics of courseware built on stand-alone machines. To explore this capability further a WWW-based Podiatry test was developed and evaluated.

In consultation with a lecturer in the Podiatry School at Queensland University of Technology a Multimedia Podiatry Test (http://www.dstc.edu.au/AU/staff/richard-taylor/podiatry/Podtest.html) was constructed on its home page. The questions for the test were taken from existing paper-based questions. Once completed, the test was demonstrated to a lecturer at the Podiatry School, and then to a practising podiatrist. The feedback from both parties was very encouraging. The lecturer considered that the functionality of the test would be a useful addition to the pedagogical tools available to him; the practising podiatrist was interested in the prospect of having available an on-line multimedia test incorporating high quality colour images and video since practising podiatrists in Australia are often spread over very large areas at considerable distance from a Podiatry School.

A Demonstration of the Educational Potential of the WWW - A Case Study in Distance Learning and Supervision

One obvious role for the WWW, using all of the above scenarios is in Distance Education [Ibrahim 1994]. The United Kingdom's main distance educator, its largest University and also its largest publisher, namely the Open University (http://www.open.ac.uk), is now beginning to exploit the WWW. Compared to a traditional distance education system of paper and post, some of the benefits of the WWW for dissemination of material include the ability of the training centres to distribute the knowledge on a large scale almost instantaneously; the reduction of mailing costs which allows distribution of pages without the overhead associated to printing costs and transport, the correction and updating of all information for all users from just one server site, the availability of a variety of different teaching styles and modes of communication between teachers and learners, facilitation of collaborative writing between authors, and improved mechanisms for students to give and receive feedback more easily.

The above Podiatry project itself was an exercise in distance education, as the research and practical work were completed in Australia with the supervisor based in the United Kingdom. All supervision of the project was performed over the Internet, with the majority of the communication using e-mail. 'Interactive' progress meetings were conducted using the simple text based TALK program, accomplished by connecting to an Internet server at a mutually convenient time pre-arranged by e-mail. When it worked well, this facility was invaluable. However, on numerous occasions the server and Internet connection suffered severe performance problems. This manifested itself by either hanging the program or slowing it down to such an extent that it ceased to be interactive and had to be abandoned at that time. Real-time meetings were also conducted using Internet Relay Chat (IRC). This helped alleviate the performance bottleneck and also provided a slightly more user friendly graphical user interface. The practical components of the project and the draft and final project reports were exchanged by e-mailing the URL address of the relevant work, and using a portable document format for the transfer.

On the one hand there was some frustration in getting the tools and services to work and with the poor quality of some of the services, but on the other hand the use of email for the bulk of the conversation exchange proved quite adequate, especially when it could be supplemented with IRC. Internet Phone took this one step nearer to face-to-face contact by allowing interactive audio. In addition, the use of the WWW enabled immediate viewing of the constructed web-pages of educational sites and of the Podiatry test demonstration. Whilst the setting up and use of these communication systems created an additional time overhead in the project that would not have been a factor in the traditional case, without them supervision of the project would not have been not have been practicable within the available time scale.

Conclusions

Assuming that the future of the WWW is secure, at least for the foreseeable future, it is inevitable that education will stand to benefit as a result of the continuing growth and development of this information-rich environment. The most significant impact is likely to come in the distance education arena where the remote use of advanced teaching materials will reduce costs and enable more students to gain a useful education. The WWW's potential collaborative features will, as they evolve further, offer greater interaction between distance education students.

The practical components of this project demonstrated the large quantity and varying quality of educational material already available on WWW. At this relatively early stage of development of the WWW it is not surprising that a proportion of the published material is not of high quality. However, it is encouraging that many of the visited educational WWW sites were useful and informative. The cataloguing of the enormous variety of educational resources available on-line, and the publishing of these lists is a useful and important activity. The continuing compilation of related WWW sites will increase the usability of the WWW in all areas, including education.

The WWW seems particularly well suited to education in a medical area such as Podiatry which is visually intensive. The WWW's inherent multimedia capabilities are well used in such an application, suggesting that its adoption for this kind of teaching environment is a real possibility. The use of the WWW and the Internet for the supervision of this project offered some valuable insights into the problems that may be encountered in the development of WWW-based education. Despite the initial time overheads, it is likely that the increase in demand for distance learning which has already begun may well be satisfied, at least in part, by the types of the WWW functionality discussed in this paper. Furthermore, the lack of face-to-face communication may be ameliorated in the future by the introduction of affordable, Internet-based, video-conferencing systems.

Above all, institutional access to the Internet and the WWW in particular must increase dramatically. Governments must propose and support schemes such as the UK Labour Party's promise (http://www.poptel.org.uk/Labour-Party/execsumm.html) to ensure a WWW facility in every school. Universities must realise the potential of the technology and invest in the infrastructure to offer WWW facilities to every faculty, again with the required support, and a recognition that the WWW can have a major impact on undergraduate and graduate education. Further development of the cable and telephone infrastructure, and the lowering of call charges will increase the availability of access still further, and as more participants contribute to the WWW the more useful it will become.

References

[Blumberg 1994] Blumberg, R.B. (1994) *An Electronic Science/Math/History Resource for the WWW*, URL,http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Educ/blumberg.mendelweb/MendelWeb94.blumberg.html

[Cho 1994] Cho, P. (1994) *IntelliTeaching - High Impact Learning*, URL, http://www.ncsa.uiuc.edu/SDG/IT94 /Proceedings/MedEd/cho/ncsa-cho.html

[Curtis & Nichols 1993] Curtis, P. & Nichols, D. (1993). *MUD's Grow UP: Social Virtual Reality in the Real World.* URL, http://lucien.berkeley.edu/MOO/MUDsGrowUp.ps

[Ibrahim 1994] Ibrahim, B. (1994) Distance Learning with the World Wide Web. *The Proceedings of the International Conference on Open and Distance Learning - Critical Success Factors, Geneva, 10-12 October 1994*, pp 255-265

[Ibrahim & Franklin 1995] Ibrahim, B. & Franklin, S.D. (1995) Advanced Educational Uses of the World-Wide Web. In *Proceedings of Third World Wide Web Conference - WWW'95, Darmstadt, Germany*, April 10-14 1995, vol 27, pp 871--877.

[Laurillard 1995] Laurillard, D. (1995) Multimedia and the Changing Experience of the Learner. *In British Journal of Educational Technology*, Vol 26, No 3, pp 179-189

[Perron 1994] Perron, D. (1994) *Learning on the WWW: A Case Study*. URL, http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Educ/perron/perron.html

[Russell & Baird 1995] Russell, D.B. & Baird, J.H. (1995) *WWW, Researchers and Research Services.* URL, http://www.scu.edu.au/ausweb95/papers/index.html#RTFToC65

[Shank & Cleary 1994] Shank, R. & Cleary, C. (1994) *Engines for Education.* URL, http://www.ils.nwu.edu/~e_for_e/

[Treloar 1995] Treloar, A. (1995) *Electronic Scholarly Publishing and the World Wide Web*, URL, http://www.scu.edu.au/ausweb95/papers/index.html#RTFToC55

Mapping the Web: A Media Typology for Information Traffic Patterns on the Internet Highway

Jens F. Jensen
Assistant professor
Aalborg University, Department of Communication, Denmark
e-mail: jensf@hum.auc.dk

Abstract: The purpose of this paper is to describe and discuss computer networks as media, with special reference to the Internet and various Internet services. The paper starts with a short introduction to the computer and computer networks seen from a media perspective. This is followed by the presentation of a new typology for information traffic patterns. Finally, the Internet and various Internet services will be examined and discussed within the framework of the established typology.

Theory: The 'Media Turn'

Computer technology does not lend itself readily to definition. It is made up of many strands and trends, like a puzzle picture that changes shape according to the point of view adopted. During its short but turbulent history of only 50 years, the computer has been conceptualized and theorized about from a series of very different metaphors and concepts. These range from a machine, a system, a tool, a calculator, a typewriter, and a dialog partnerñto an artificial intelligence, a brain and an expert.

The central idea of this paper is to explore computers and computer networks from quite another point of viewñfrom a media perspective, i.e. as intermediate technological agencies that permit communication by transmitting signs from human senders to human receivers and as such are used for the transmission of information, conversations, requests, entertainment, education, expression of emotional experiences, and so on. Therefore, in this case, the analogies and metaphors that will be used to describe and come to terms with computers and networks are not drawn from the domains of machines, tools, human brains etc., but from the realm of media, e.g. newspapers, books, telephones, television and radio [see Andersen, Holmqvist & Jensen 1993].

There are several reasons why the media perspective has become more and more relevant, even urgent, in relation to computers and computer networks. Only the most important of these reasons will be suggested in the following. First, computer technology and the way that computers are used has changed from a primary function as a machine which processes data, toward a primary function as a machine which disseminates information, i.e. from tool to media. This shift in emphasis is significantly reflected by concepts such as: multi*media*, hyper*media*, computer *media*, computer mediated communication, computer networks, etc., all of which have, during a very brief time, won wide acceptance within the field of computing, the world of advertising, everyday usage and in scientific discussions. Secondñand as a related developmentñmore and more computers are being connected to networks which emphasize the communicative aspects, just as computers are taking over a long list of communicative functions previously performed by other media, and in that way are beginning to play an increasingly important role in human communication. Third, a long list of older media (TV, radio, telephone, fax etc.) are now being digitized, so that their technical architecture, for all practical purposes, becomes similar to the computer. The result is that these media are converging with each otherñand all of them are converging with the computer. And fourth, the ever increasing complexity of our society necessitates more flexible and intelligent ways of dealing with information and communication, where today computers offer the most advanced techniques for handling information. Computer technology is thus on the way to establishing itself as the dominating social technology for communication, storage, processing and production of data, information, and meaning. This has already happened to such a degree that the computer has been dubbed 'the fourth media', i.e. the new media which will take over the position of, respectively, newspapers, radio and TV as the socially dominant media. Taken as a whole, these developments indicate the necessity for scientific research on the computer as media, so that we may speak of a 'media turn' in informatics and computer science [see Jensen 1993].

The extensive penetration of computer and network technology in human information and communication systems and the extreme complexity of computers as media necessitates, among other things, a typology or classification of information services and communication, that can provide us with a general view of the field of computer mediated communication. There have been numerous attempts to categorize and typologize communication and media technologies, including computer technologies. The most widespread classifications have been based on technical properties, forms of presentation, or the information content. But as different media technologies merge in digitization, different forms of presentation merge in multimedia, and different content merges in multi-funtional or full service networks, these types of current classifications seem less and less instructive and satisfying.

The Classification of Information Traffic Patterns

One possible alternative way of classifying new media technologies, which overcomes some of these drawbacks, is the media typology developed by [Bordewijk & Kaam 1986]. A media typology which they have designated as »a new classification of tele-information services«, *de facto* motivated by the present explosion of tele-information systems, to be understood here as the merger of digital telecommunication and computer technology. The distinctive mark of this typology is that it is defined independently of the technical design of the media, the form of presentation, and the content of information, and instead based on social power relations and power positions, which constitute different 'information traffic patterns'. One of the most widely used metaphors for computer networks as media has been the 'information Super Highway'. (The extent of its dissemination is among other things reflected by its many international variations: Der Infobahn (de), Autoroute de l'information (fr), Autostrade dell'informazione (it), Autopistas de la informacion (es), Autoestrada de informação (pt), informations-motorveje (dk).) A metaphor that was already a cliché before the majority understood its implications, and a metaphor that is still heatedly debated, it has been heavily criticized for creating wrong conceptions. In this particular context, however, the metaphorñas already indicated by the title of this paperñis appropriate, since it relates directly to the mapping of information traffic on the Internet Highway, which Bordewijk and Kaam refer to as 'the idealized information traffic patterns'. In the following, I will give a brief description of the different information traffic patterns, primarily based on Bordewijk and Kaam's presentation.

The typology takes its point of departure in two basic questions, partly concerning the central or decentral position of sender and receiver, and partly concerning the ownership of the information and the control of the access to and the use of the information. Here (C) is the information service provider, and (i) is the information service consumer. These terms have the double advantage of including technical facilities, but leaving the direction of the information flow open. The two questions can be stated as follows:

_ Is the transmitted information owned by an information service providing center (C) or an individual information service consumer (i)?

_ Is the transmission and use of the information controlled by an information service providing center (C) or an individual information service consumer (i)?

The answers to these two questions can be combined in a single matrix definition, thus giving four possible combinations of answers, representing what Bordewijk and Kaam describe as »four idealized information traffic patterns«: transmission, conversation, consultation, and registration [see fig.1].

Top-left: Transmission. If the information service center produces and is the owner of the information, as well as the sole controller of the choice of and time for distribution of the information to the information consumer, we have a case of *transmission*. (Bordewijk and Kaam use the term 'allocution', but for the sake of consistency in the terminology I have chosen the term 'transmission'). In this traffic pattern the flow of information will thus exclusively run in one direction, from the service center (C) to the individual consumer (i) as illustrated in [Fig. 2]. If the information center serves several consumers, which is of course often the case, we have the generalized pattern noted in [Fig. 3], in which (C) is the service center and (i1, i2, i3 etc.) are individual consumers. In other words, this is the kind of communication often referred to as one-way communication, one-to-many communication, mass-communication or mass media.

Characteristic features of the transmitting media are that a large (potentially infinite) amount of information is available to the information center; that all consumers in principle receive exactly the same information and often receive the information synchronously; and that, with respect to power, the media are strongly asymmetrical and centralistic, since the center has maximum power over the information traffic by deciding content as well as programming, and the consumer has minimal power, being totally subject to the decisions of the center in both aspects. In practical situations, consumers will, more often than not, have the possibility of giving some kind of feedback, but because the model has been designed as an 'idealized' information traffic pattern, this aspect has been left out. The prototypical examples of the transmission pattern are of course classical broadcast media such as TV and radio. If we refrain from the requirement for synchronous or simultaneous distribution and reception we could also include media such as film, books, newspapers etc.

Bottom-right: Conversation. If the individual information service consumers produce and own the information, and if control of the means of distributing and handling the information are divided equally between those consumers (at least in this idealized case) the resulting situation is diametrically opposed to the one above, and we have a case of *conversation*. In this instance, the information flow runs in both directions, between the two individual consumers, and there is an equal exchange of information as illustrated in [Fig. 4]. In a more general model, the connection between the two information consumers (i1 and i2) will often be provided by an information service center (C), but the center represents a purely technical facility, neither intervening in the production of information or in the time of distribution, just as the center often serves several consumers and handles numerous conversation-connections, as noted in the generalized pattern of [Fig. 5]. This is the kind of communication often referred to as one-to-one communication (or many-to-many communication), dialog, conversation etc.

Characteristic features of the conversation pattern are therefore: that the information, as well as the time for exchange of information, are totally controlled by the consumers; that the distribution of power (at least in the idealized form) is completely symmetrical and decentralistic (contrary to the above mentioned transmission pattern); and that the traditional distinction between senders and receivers therefore tends to dissolve. The prototypical example of the conversation pattern is of course the telephone, where the individual consumer decides whom to call, when to call, and what to talk about, whereas the telephone system plays a purely technical and mediating role in routing the telephone calls between the consumers. If we again refrain from requiring a synchronous mode and instantaneous interaction, other examples of conversation could be: telex, fax, ordinary mail etc.

Until now, communication and media studies have primarily based their models and insights on the two above mentioned communication patterns: the transmitting pattern in mass communication studies, and the conversation pattern in interpersonal communication studies, whereas the two following communication patterns have been markedly underexposed.

Bottom-left: Consultation. If the information is produced and owned by a central information service center (C), but the individual information service consumer (i) controls which information is to be delivered and when to deliver it, we have a case of *consultation*. In this instance, the information service center only delivers information on request from the individual consumer (thin line), and reversely the individual consumer only gets the information required, at the time required (thick line) as illustrated in [Fig. 6]. In this pattern too, the center can normally be consulted by, and serve, several information consumers, resulting in the generalized pattern [Fig. 7].

The consultation pattern represents markedly different power relations in comparison with the two patterns considered above. Characteristic features of consultative media are that they have an asymmetric, but distributed division of power, where the power over the content is held by the center, and the power over the distribution held by the consumer; that theyñcompared to the transmitting modeñrequire a higher degree of activity or interactivity on behalf of the consumer; but that they in return offer far more freedom and flexibility in the choice of information, making more individual and selective uses possible. Exemplary representatives of the consultative pattern are found in various forms of 'electronic memories' such as CD-ROMs, harddiscs, data bases, online services, etc., in various forms of printed media like reference works, dictionaries, encyclopedias. It is even possible to argue that newspapers, magazines, periodicals and books belong in this category, to the extent that they are considered collections of information units which can be selected and read by the information consumer, at a time that is convenient. On a more general level, you could also say that institutions like libraries, bookstores, etc. follow the consultation mode.

Top-right: Registration. And finally, the opposite situation: when the information is produced by an individual information service consumer (i), but the use of the information is handled and controlled by an information service center (C), we have a case of *registration*. Here the task of the center is no longer to issue information, but to collect it. This information pattern has two variations. If the information is delivered by the individual on request from the center we have an inversion of the information flow of the consultation pattern ([Fig. 6] and [Fig. 7]) as illustrated by the generalized pattern in [Fig. 8]. If, however, the information is collected from the consumer without a request from the center, we have an inversion of the information flow of the transmitting model, where the center no longer sends information to, but collects information from, the consumer ([Fig. 2] and [Fig. 3]) as depicted by the generalized pattern in [Fig. 9]. The collected information can in turn be processed, computed, (re)arranged, etc. by the center. This kind of pattern could perhaps be referred to as many-to-one communication.

The registration pattern is thus also characterized by an asymmetric, but distributed division of power, only in reverse order from the transmitting and the consulting patterns, since the information is produced and delivered by the consumers, whereas the use of the information, i.e. the adoption, interpretation and distribution (concealment, publication etc.) is controlled by the central information service. Examples of the registration pattern (in the reversed consultation version, i.e. 'on request') would be (tele-)opinion polling, TV-viewers voting by letter or phone, shopping systems, etc., or (in the reversed transmission version, i.e. 'without request') tele-alarm, video surveillance, logging of computer systems, electronic surveillance, security and alarm systems, etc. But other examples, on a more general level, could also be well known services such as news agencies, civil registration, tax authorities, and other public (accessible as well as non-accessible) registration systems, etc.

Commercial/public service. In addition to Bordewijk and Kaam's two questions, I now wish to introduce a third concerning the overall interest behind the transmission of information. This is a dimension of the info-traffic which seems to be of increasing relevance as the Internet becomes more and more commercialized. The question can be stated as follows:

* Is the transmission of information motivated primarily in the interest of the information service providing center (C), i.e. a commercial service, or in the interest of the information service consumers (i), i.e. so-called *public service*?

This question serves as the basis for the construction of a third dimension in the matrix, thus establishing eight possible patterns or media types of either commercial or non-commercial/public service: transmission, conversation, consultation and registration [see Fig. 10].

Information Traffic Patterns on the Internet Highway

As is apparent from the above description, the eight information patterns represent very different forms of communication and social power relations. It is now the thesis of this paper that the current media landscape in general, and the services and applications on the Internet in particular, can be categorized in terms of these information patterns. The remainder of the paper will elaborate on this thesis.

If you consider the current new media developments that have arisen in connection with digitization, telematics, computer media etc., within the framework of the two dimensional matrix, they can best be described as a relative movement away from the *top-left* position, towards the other positions, i.e. from the traditional transmission pattern, towards the conversation, the registration and especially the consultation patterns. This evidently also implies a general movement away from the strongly asymmetrical, centralistic power structure, towards a larger symmetry in the distribution of power. In connection with a discussion of Bordewijk and Kaam's matrix typology [McQuail 1987] writes: »The new media seem to offer the potential of a shift on the balance of power away from the sender and towards the receiver, making much more content of all kinds accessible to users and choosers without dependence on the mediating and controlling systems of mass communication«. And he concludes: »This seems to indicate a general increase in individual freedom to gain information and a reduction in the dominance of centralized sources«. While this can undoubtedly be said to be correct from a general perspective, both for the collective new media landscape and for the Internet, there are still considerable modifications and opposing indications.

Since its historic beginning as the military ARPA-net, Internet has primarily functioned as a communication network. The entire structure built into the protocols and hardware that combine to constitute the Internet has been designed for point-to-point communication, and even a very non-hierarchical, decentral communication network. As such, from its start, the Internet was conceived within a conversational pattern and this type of traffic is still considerable. The dominant application in this pattern is obviously e-mail, understood as electronic post, which makes it possible to send written messages from one consumer with an electronic address to another (or several others). In this case, the center is the packet switching transmission system in which data are transported from different origins to different destinations. In spite of its asynchronous features, e-mail can thus be characterized as a relatively typical example of conversational media. But also Usenet and some mailing lists (especially when the individual members are active) can be seen as conversational patterns in the form of many-to-many conversations. Synchronous or quasi-synchronous conversational patterns can be found in various chat programs, in net-based online games such as MUD, MUSE, and MUSH, and in various forms of computer based video-telephones made possible by new broadband, switched, two-way networks. For obvious reasons, most information exchanged in this pattern is of a non-commercial character, although examples of commercial conversation can be found in online games, consultation and help services, etc.

The other dominant traffic pattern on the Internet is undoubtedly consultation. As is clear from the above description, digital media's large storage capacity and facilities for quick retrieval of information, are special strengths in relation to the consultative pattern. A large portion of the resources, services and tools on the Internet can also be more or less understood within the consultative pattern. A brief list follows. TELNET, understood as a service or an application that makes it possible for individual consumers to select and connect to remote hosts on the Internet and interactively use the services and programs they offer so that the consumer's computer actually becomes a terminal for the selected host computer. Anonymous FTP (File Transfer Protocol), understood as a service or an application that makes it possible for the individual user to select and, for example, copy files over the Internet from a host or file server to his own computer. Gopher, understood as an information search service or program that, using a menu hierarchy, makes it possible for the individual user to navigate through, find, and retrieve information on the Internet. World Wide Web, understood as a global network of interconnected documents which allows the individual information consumer, using the principle of hypertext links, to jump from one to another. And there are of course various other forms of online information and data base services. This pattern can also be found in a public service version (e.g. university libraries) which have predominated, as well as a commercial version (e.g. large media conglomerates, archives) which now appear to be taking over the dominant role.

But the consumer oriented traffic patterns are not the only ones that influence the Internet. The registration pattern is f.ex. also represented, although in a somewhat less dominant (or less visible) way. In its 'on request' version it can be seen in shopping systems, ticketing, registration procedures, financial transactions, various forms of Internet voting, etc. And in the 'without request' version, it can be seen in: the logging of information traffic, the counting of homepage visitors, as well as other forms of collection and processing of the electronic fingerprints left behind while moving about the Internet. Once again there is both a public service version (Usenet polling, etc.) and a commercial version (shopping systems, etc.)

For some time now, it has been the general impression, and also a point made ideologically, that computer networks, and the Internet in particular, represent the absolute opposite of the transmitting pattern. It has been claimed that the Internet is not a broadcast media, is not a one-to-many form of communication, but that it is, on the contrary, designed as a many-to-many form of communication where it is possible for everyone to express themselves, i.e. a conversational pattern. However, it does not appear to be quite that simple. [Steinberg 1996] among others, has recently pointed out that as »the Internet grows to support new services beyond e-mail and file transfer, its antipathy to broadcast is starting to seem more like a bug than a feature«. If one were to set up an Internet service from Copenhagen today, which delivered updates of a certain type of information at regular intervals, and there were 100 customers in California, 100 identical copies of the message would be sent around the world over the Internet. Here is a situation where it would obviously be much more economical and net-efficient to send one message to, say, San Francisco and from there send it in 100 directions. »What's needed is some way to do one-to-many communication over the Net« concludes Steinberg.

This type of information traffic is beginning to be found on the Internet f.ex. via the so-called Multicast Backbone (MBone). The idea behind the Mbone is that Internet's infrastructure and protocols are used to transport data but with the help of various supplemental algorithms that eliminate the redundant data transmission. The goal of multicasting is thus to minimize network traffic and thereby burden the Internet as little as possible. This approach is naturally becoming more important as the Internet becomes more and more overloaded with traffic. Via multicasting it is thus possible to send data to supporting sites and thereby, for example, transmit live broadcasts of technical conferences, real time radio, TV or multimedia applications to thousands of computers all over the world. In Steinberg's words: »Let's just hope that the Net can be made to support broadcasting without actually becoming a broadcast medium« [Steinberg 1996]. It should be noted, however, that besides transmission the Mbone is capable of supporting interactive two-way services, i.e. the consultation, the conversation as well as the registration patterns. Some odd examples of the transmitting pattern are the famous WWW homepage, The Trojan Room Coffee Machine, which constantly updates pictures of the coffee machine at the University of Cambridge computer laboratory, and the WWW homepage The San Francisco City Camera, that transmits a similarly updated panorama picture of San Francisco from a camera placed on top of the Fairmont Hotel. The same pattern is also apparent, aside from the synchronous aspect, in mailing lists and subscriptions to newsgroups where a number of individual consumers are sent the same amount of information, especially when a large number of so-called lurkers are involved. Like TV and radio, the transmitting pattern can also be found in both a public service (scientific conferences, lectures, etc.) and a commercial version (entertainment etc.). See [Fig. 11] for an illustration of the eight information traffic patterns within the framework of the three dimensional matrix.

It should now be obvious that the Internet, when considered from a general point of view, as *one* unified medium, is not exclusively in a single matrix position. On the contrary, it assumes the character of what [Bordewijk & Kaam 1986] describe as »multipattern-services«, i.e. services where »several patterns occur … simultaneously«. The Internet, and computer networks in general, seem to share this 'multipatterned' feature with a number of other contemporary information services or systems such as telephones and television. From originally being a prototype for the idealized conversation pattern, the telephone and its specialized services now also provide consultation (e.g. weather forecasts), registration (e.g. settling of accounts) and even transmission patterns (e.g. live transmissions of debates from the parliament). Television, which was originally a prototype for the idealized transmitting pattern, can now be found in its advanced form as interactive television which also provides consultation (video-on-demand), registration (wagering), and even conversation patterns (video-phone) (see [Jensen 1996a] and [Jensen 1996b]). It is, nevertheless, reasonable to maintain that the Internet, in terms of information traffic patterns, can be considered quite a unique medium in that it undoubtedly constitutes the most complex 'multipattern service' in the contemporary media landscape.

Even if we focus on the individual Internet services or applications, the boundaries between information patterns are not necessarily razor sharp. A given homepage on the World Wide Web can present both consultative characteristics (homepages are selected for visits by the individual consumer), registration characteristics (counting of visitors), conversational characteristics (forms for direct e-mail contact with the homepage owner), and possibly transmission characteristics (if the homepage is constantly updated such as the Trojan Room Coffee Machine). In this case too, the specific services and applications are best described as 'multipattern services', where the idealized information traffic patterns are only building blocks.

From this point of view, as 'multipattern services', the matrix opposes itself to a certain extent. The very construction of the matrix and its application to current media phenomena like the Internet, demonstrates that the characteristic patterns and the sharp distinctions disintegrate in favor of more complex and hybrid information patterns and media forms. The typology, nevertheless, offers a suitable description tool, in that the different new media and services can be significantly described as a combination of several information patterns.

There are several further points to be made from this matrix typology. For reasons of space, however, I will only be able to give a few brief indications (see [Jensen 1996c] for a more complete study). If we consider education within the framework of the two dimensional matrix it appears that all four traffic patterns are represented, although on a different scale: the transmission mode when the teacher is lecturing, the consultation mode when the pupils study their books or ask questions, the registration mode when the teacher subjects pupils to examination, and the conversation mode when pupils (and the teacher) discuss the subject matter. It can be inferred from this, among other things, that systems for distance learning must provide possibilities for all four information patterns. This is one of the reasons why, contrary to expectation, education through transmitting media like TV always has been a failure, and conversely, why multi-pattern networks now offer a more promising possibility [see Bordewijk & Kaam 1986]. Similar considerations can be made in relation to other spheres of application.

Regarding legal issues, it can be said that because the eight patterns refer to eight mutually exclusive social power relations, similar regulatory aspects can be considered relevant for all services that follow one and the same information traffic pattern. The three dimensional pattern matrix can thus serve as a basis for drawing up a legal framework: the left column of the matrix emphasizes copyright issues, responsible press laws, censorship, content regulation (e.g. the Communications Decency Act); the right column emphasizes civil liberties, freedom of speech, security, protection of privacy, including issues such as encryption, secrecy of the mails; the top column emphasizes public access and control procedures; the bottom column emphasizes access to public information, the free flow of information versus state interference; the front column emphasizes market regulation and regulation of monopolization; and the back column emphasizes issues relating to public service. Each information pattern thus lies at the intersection of three lines of regulatory concerns. In this way the matrix provides us with a map of the social and legal position of new information services and computer networks (which makes it clear that the 'one-size-fits-all' standard will not do in this case). In similar ways several other inferences concerning regulatory aspects can be drawn from the matrix.

The Internet is undoubtedly on its way to becoming the new socially influential medium, and a mediumñor a complex of mediañof a hitherto unprecedented complexity, flexibility and power. I believe that the theoretical approaches and frameworks presented here are relevant for analyzing (and designing) computer networks, services and applications, and that their relevance will increase in the years to come. And I believe that they will particularly prove indispensable for the understanding of the Internet as mediañand hence for the conceptual mapping of the Web.

References

[Andersen, Holmqvist & Jensen 1993] Andersen, P. B., Holmqvist B. & Jensen, J. F. (eds.) (1993). *The Computer as medium*. Cambridge: Cambridge University Press.

[Bordewijk & Kaam 1986] Bordewijk, J. L. & van Kaam, B. (1986). Towards a new classification of Tele-Information services. *Intermedia*, 14 (1).

[Jantzen & Jensen 1993] Jantzen, G. & Jensen, J. F. (1993). Powerplay. Power, Violence and Gender in Video Games. *AI & SOCIETY. The Journal of Human-Centred Systems and Machine Intelligence*, 7 (4), 368-385.

[Jensen 1993] Jensen, J. F. (1993). Computer Culture: The meaning of technology and the technology of meaning. A triadic essay on the semiotics of technology, in [Andersen, Holmqvist & Jensen 1993]

[Jensen 1995] Jensen, J. F. (1995). *Multimedier og Teknologiudvikling. Rapport udarbejdet for Statsministeriets Medieudvalg* 1995. (MultiMedia and technological development. A Report to the Media Commission under the danish Department of State), Copenhagen: Medieudvalget, Statsministeriet.

[Jensen 1996a] Jensen, J. F. (1996). Interactive Television: 'Coming soon at a screen near you'. *K&K*, 80, 27-66.

[Jensen 1996b] Jensen, J. F. (1996). Interactive Television. Working paper from the research program: The Aesthetics of Television. Aarhus: Department of Information and Media Science.

[Jensen 1996c] Jensen, J. F. (1996). Mapping the Internet Highway. A media typology for Information Traffic Patterns on the Internet. *Journal of Universal Computer Science* (forthcoming).

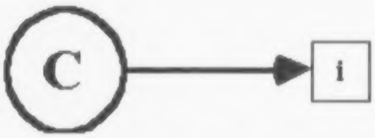
[McQuail 1987] McQuail, D. (1987). *Mass Communication Theory. An Introduction*. London: Sage.

[Steinberg 1996] Steinberg, S. G. (1996). Multicasting. The Internet's connected to the MBone. *Wired*, 4 (6), 86.

Appendix: Figures

	Information issue by center	Information issue by consumer
Programme control by center	**TRANSMISSION**	**REGISTRATION**
Programme control by consumer	**CONSULTATION**	**CONVERSATION**

Fig. 1: Pattern matrix for the four 'idealized information traffic patterns'



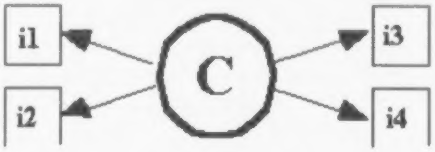
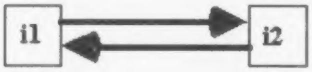


Fig. 2: Transmission: information flow **Fig. 3:** Transmission, generalized pattern: Information

from service centre (C) to individual consumer (i) flow from service center (C) to individual consumers (i1, i2, i3, etc.)



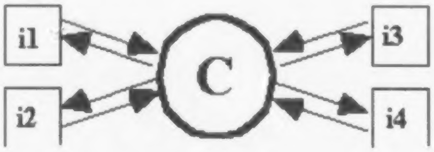
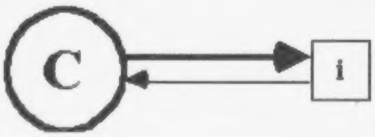


Fig. 4: Conversation: Information flow **Fig. 5:** Conversation: generalized pattern

between individual consumers (i1, i2)



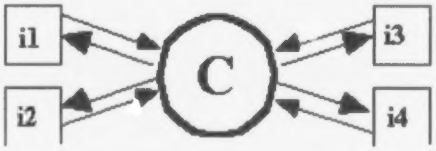
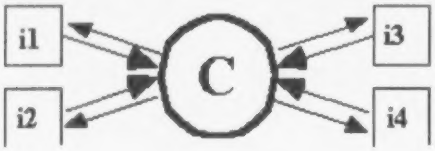


Fig. 6: Consultation: information on request **Fig. 7:** Consultation: generalized pattern (request: thin

(request: thin line) line)



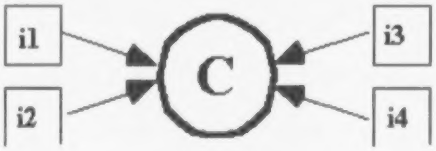


Fig. 8: Registration: generalized pattern **Fig. 9:** Registration, generalized pattern, without request

(request: thin line)

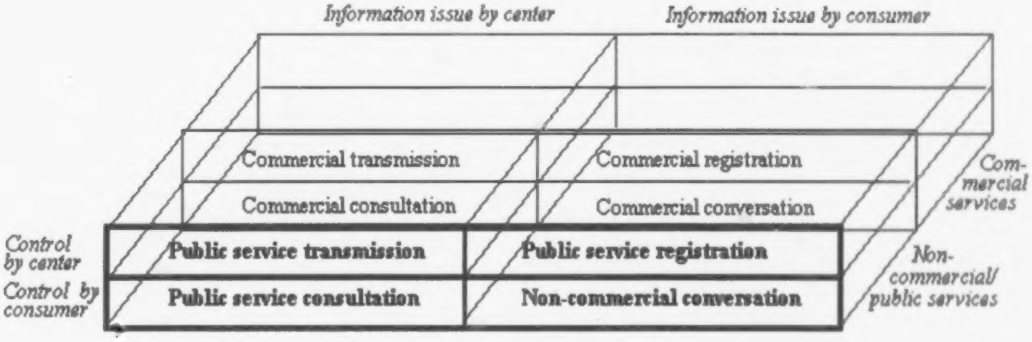

Information issue by center
Information issue by consumer
Commercial transmission
Commercial registration
Commercial consultation
Commercial conversation
Com-mercial services
Control by center
Control by consumer
Public service transmission
Public service registration
Public service consultation
Non-commercial conversation
Non-commercial/ public services

Fig. 10: 3-dimensional pattern matrix for the eight 'information traffic patterns'

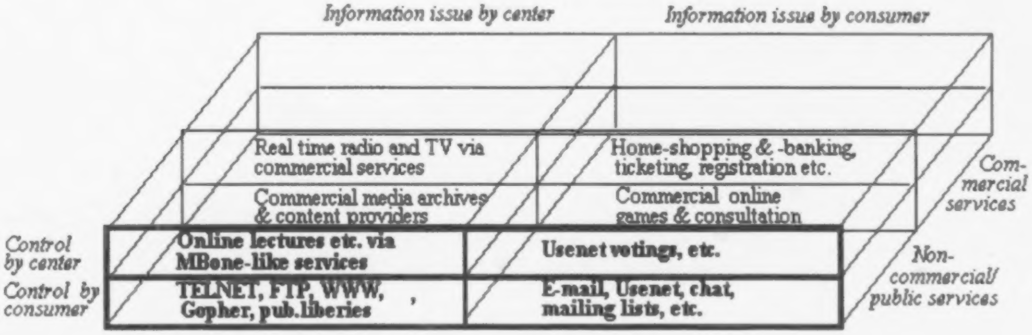

Information issue by center
Information issue by consumer
Real time radio and TV via commercial services
Home-shopping & -banking, ticketing, registration etc.
Commercial media archives & content providers
Commercial online games & consultation
Com-mercial services
Control by center
Control by consumer
Online lectures etc. via MBone-like services
Usenet votings, etc.
TELNET, FTP, WWW, Gopher, pub.liberies
E-mail, Usenet, chat, mailing lists, etc.
Non-commercial/ public services

Fig. 11: Illustration of the eight information traffic patterns on the Intnernet

BEST COPY AVAILABLE

Automated Management and Delivery of Distance Courseware

W. Lewis Johnson, Tyler Blake, and Erin Shaw University of Southern California, U.S.A.
andes@isi.edu

Abstract: This paper describes a system called ANDES for management and delivery of distance education courses. ANDES enables students to study at home, at their own pace, as well as interact with instructors and other students in virtual "classrooms." It uses World Wide Web technology for transmission and delivery, with extensions relevant to distance education. It seamlessly integrates course materials on CD-ROM with materials that have been downloaded over the Internet. Student progress through the course is tracked automatically for evaluation purposes. A high-level authoring language supports rapid development of courseware materials. Custom interfaces for course instructors and administrators provide online views of student work and progress.

1. Background

As the World Wide Web has developed, it has attracted considerable attention as a potential vehicle for courseware delivery. Universities with active distance education programs, such as The Open University [Open 96] and the University of Athabasca [Athabasca 96] are making increasing use of Web. Although the Web usually plays a minor role as an adjunct to conventional materials such as printed textbooks and study guides, there have been experiments using the Web as the primary delivery mechanism, and HTML as the authoring language [Holt et al. 95].

The Web has a number of limitations that stand in the way of effective use for distance education. Accessing pages can be slow, especially by modem; this limits the rate at which students can navigate through material, and effectively precludes the use of large multimedia assets such as video clips. HTML pages are static presentations; making them interactive requires expertise in CGI or Java programming, and is therefore beyond the reach of the typical courseware author. Maintaining networks of Web pages is an error-prone process, as is evidenced by the large number of obsolete links on the Web. HTML is a general-purpose authoring language, and does not commit to any particular look and feel. It therefore provides no built-in support to authors who wish to maintain a consistent look and feel throughout their courseware, or who wish to define templates for common courseware elements. Finally, it is difficult to obtain detailed statistics about student progress through the course material; the log files that Web servers maintain give only a partial picture of student activity.

Research efforts are underway to address some of these limitations. For example, the Hyper-G multimedia system maintains document links in a separate database, which helps prevent links from going stale as documents are modified [Maurer&Lennon 95]. The Intelligent Bandwidth project is developing ways of caching Web pages throughout the network, in order to improve effective bandwidth [Touch 96]. However, these general solutions do not adequately meet the specific needs of distance education. ANDES, in contrast, is designed specifically to support distance education courseware; it achieves dramatic increases in effective bandwidth, and provides specialized capabilities such as authoring support and student tracking. Multiple applications are combined in a single environment, avoiding disruptive shifts in students' attention. These goals are accomplished while adhering to standard communication protocols and employing common Web software.

2. The Structure of ANDES Courses

The School of Cinema-Television at USC is developing the initial courses to be delivered using the ANDES system. Courseware content is delivered via video-taped lectures and CD-ROM-based workshops. The structure of the courses adheres to the following principles of human factors design and applied learning theory.

1. Emphasis on experiential learning -- Workshops are designed to require students to apply facts and concepts in a creative problem-solving process. For example, if a lecture describes the principles of lighting actors on a film set, a correlated workshop might present the users with a simulated set and continuous controls for light intensity and position, and ask them to produce a specific effect by manipulating the controls.

2. Optional emphasis on mastery -- The workshops can be repeated until satisfactory results are obtained, which is important since experiental learning activities yield a fair number of ``mistakes". However, ANDES can also be used to deliver courses adhering to more traditional competitive approaches.

3. Promotion of a "class" identity through group communication -- Course members are subdivided to small, manageable units (i.e. classes) and assigned a specific teaching assistant who will direct on-line discussions. The units can be created at random or based on instructionally relevant criteria. These classes provide the basis for student interchange, and contribute to the feeling of being involved in group learning activities.

4. Variable control of course sequencing -- A major advantage of remote delivery is that students can follow a self-paced, self-scheduled program. However, if some students advance much faster than others, the interchange among class members will be less effective. Each course and instructor will need to decide how to regulate advancement. The system supports everything from completely self-paced to entirely forced-paced approaches.

All courses incorporate the following student experiences, either in sequence, or with branches and iterations: 1) video presentation of lectures enhanced with broadcast quality production values, 2) interactive workshop exercises requiring application of lecture material, 3) instructor feedback via email, commented workshop screens, and chat group discussions, 4) class discussions of workshops via Internet Relay Chat and/or phone connections, 5) scheduled electronic office hours for students to ask questions of the professor and/or TA, and 6) occasional guest lectures.

ANDES courses are designed by analyzing each educational objective in a weighted consideration of four major factors: relative media effectiveness, time-criticality of the information, cost to implement and the availability of client hardware. From this analysis, the author determines to what platforms to assign the content. For example, in the pilot film course, the quality of CD-ROM-based video is inadequate to depict the production concepts to be taught; thus, the film clips are presented on videotape. Guest lectures need to be timely; thus, they are stored in reduced form (e.g., compressed audio plus still visuals) for smooth streaming to the user's machine.

ANDES Architecture

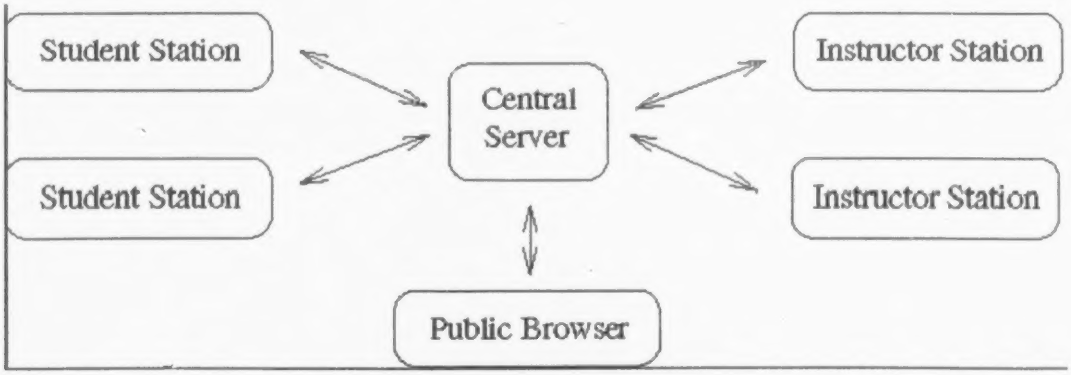


Figure 1: ANDES system configuration.

ANDES is designed to support students, teaching assistants, and course administrators. The roles of the clients and servers in the ANDES architecture are rather different from those in typical client-server architectures. Each student station is able to operate autonomously; when students are working through the course materials on their own, and are not interacting with instructors or other students, no connection to the central servers is required. Connection is established only at the beginning and end of each session, and during live class interactions.

3.1 The Student Mini-Server

Autonomy is accomplished by installing a Web server and CGI program locally on each ANDES student work station. This *student mini-server*, presents course materials to the student as requested, and tracks the student's progress through the course. The mini-server presents the course materials using a conventional Web browser, namely Netscape. The student mini-server contacts the central server automatically when needed, in order to obtain updates for course materials and report on student progress. The mini-server approach provides a number of advantages over the conventional Web as a courseware delivery platform. Storing the course materials locally on the student machine greatly improves system response, and largely eliminates download time during courseware delivery. Autonomous courseware delivery reduces the amount of Internet connect time required, and thus reduces the connect charges to Internet service providers. Because the mini-server resides on the student's machine, it can do a better job of tracking student activities and controlling the user interface than a server residing on a remote machine can. Because the mini-server pre-fetches all lesson materials at once, it does a better job of improving system response than general-purpose mechanisms such as caching.

ANDES refers to all assets by *handles*, or aliases that are resolved at runtime to yield an absolute path and file name. Handles enable courseware developers to replace or supplement assets over the life of a course. A global dictionary is used to store the actual pathnames. This approach allows courseware to be updated at any time, by downloading the updates from the central server and updating the handle dictionary. It also permits performance tuning of the courseware, e.g., copying CD-ROM assets temporarily to hard disk.

3.2 Student Workstation Organization

Fig. 2 shows the architecture of the student workstation. It includes a Web browser (Netscape), a Web server (MacHTTP or WINhttpd), and several additional processes for retrieving, displaying, and managing courseware.

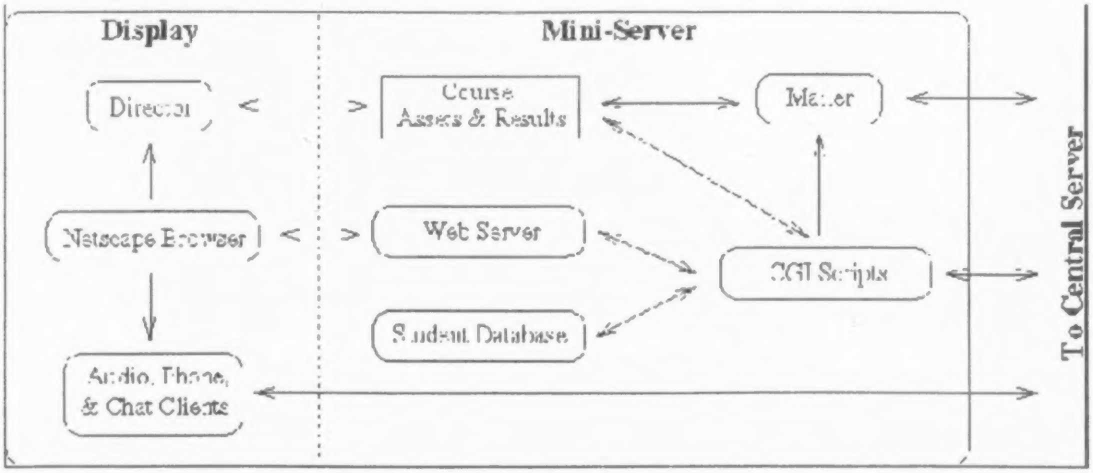


Figure 2: Student workstation architecture.

The Netscape browser communicates with a number of helper applications in order to present courseware to the student. A customized application is used to startup workshops that have been authored in Macromedia Director. The mini-server currently runs these workshops as separate, full-screen external applications because the Netscape plug-in for Director movies, Shockwave, does not yet support local file I/O. Other helper applications manage chat sessions, play audio transmissions (both live and recorded) and handle two-way voice conversations. The user interfaces of the applications are being adapted and simplified for use specifically within the context of an ANDES course.

The mini-server performs a number of functions, all implemented by scripts employing the Common Gateway Interface (CGI) protocol. It notifies the central server when a student logs in or out of the system. It locates and selects the most up-to-date versions of courseware assets for presentation. It processes high-level descriptions of course screens, written in ANDES Text Markup Language (ATML), and translates them into HTML presentations for display in the Netscape browser. It updates a database, which includes information about the course materials (when they were last updated) and about the student (how far he or she has progressed through the material). It maintains a log of student actions, both those taken while interacting with the Netscape browser and those taken while playing Director movies, and automatically transmits the log to the central server.

The student mini-server and ANDES central servers communicate using a combination of two-way Internet connections (e.g., when the student starts a lesson) and electronic mail (e.g., to transmit courseware updates). A mailer running on the student machine then filters out messages containing courseware updates, and saves the updates on the student's hard disk. Electronic mail is also used for transmitting student work to the central server. The use of electronic mail reduces the amount of Internet connect time required, and reduces the burden on the central servers. This helps to ensure that ANDES can be easily scaled to handle large course sizes.

3.3 Central Server Organization

Fig. 3 shows the organization of the ANDES central servers, which are used to support the various service requests from the student workstations. The central servers can be divided into two groups: those that manage interpersonal communication (chat, audio, and phone servers), and those that manage and distribute course materials. The focus of development in ANDES has been on the latter.

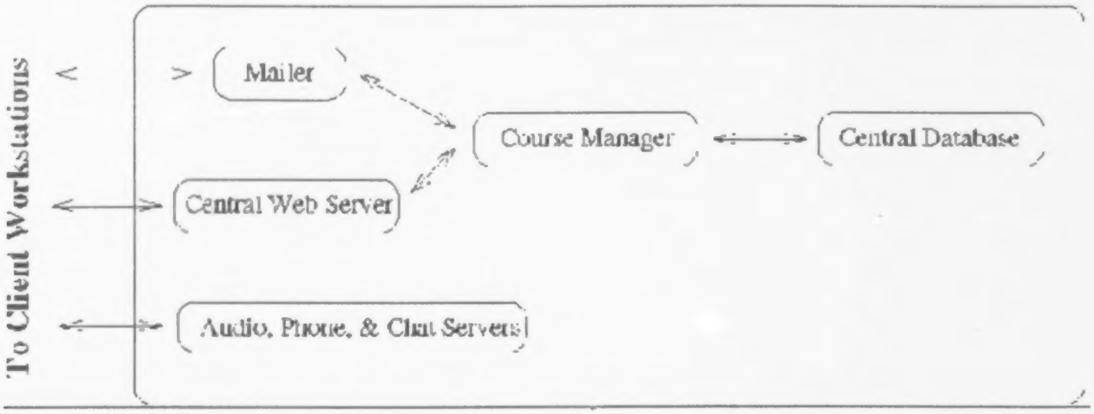


Figure 3: Central server organization.

When a student attempts to log in to the course, or when a student mini-server requests updates of course assets, the request is sent to the *course manager*. The course manager is responsible for validating the request for security purposes, determining what information is required by the client, and delivering the information. The course manager also handles distribution of courseware updates, and requests for information about a student's progress. If a lesson is updated, the course manager determines which student workstations require updates. It is sensitive to the students' progress through the course; if an instructor updates a workshop that a student has already completed, the course manager will inform the mini-server that it is necessary to repeat the workshop.

The Central Web server serves as the course manager's interface to instructors, course administrators, and the interested public. The Web server invokes a CGI stub which connects to the course manager, relays the requests, and then displays the responses on the client's Web browser. The global state of the course is maintained in an Oracle database, which records student data and courseware organization. Student exercises and logs of student activities are stored in the database as well. The databases residing on the student workstations include a subset of the data maintained in the central database.

4. Processing and Communication Details

All communication between the mini-server and course manager is handled via the ANDES protocol, which is based on the Web's Hypertext Transfer Protocol (HTTP). ANDES requests are embedded in the *path* component of HTTP Uniform Resource Locators (URLs). A request contains the script arguments and other necessary data that are passed to the servers for processing. This approach is similar to that used by other context-sensitive Web-based applications, such as I-Doc [Johnson&Erdem 96] and Dienst [Davis&Krafft 95].

The mini-server sends two kinds of requests: *login* requests and *logout*. These functions synchronize the system and guarantee that students receives the necessary course assets. If the student's computer crashes in the midst a session, before a logout request has been issued, this will be detected the next time a login is attempted.

The login connection is as brief as possible and consists of an authentication and a check for assets. If there are assets to be downloaded, the mini-server will inform the student how long the download will take. The student has the option to continue with the download or to quit and log on later at a more convenient time. The assets may either be transmitted immediately using the HTTP connection, or via electronic mail. When the download is confirmed, the login function closes the TCP connection and reconfigures the student's Web server to run locally. When the system is ready, the student is presented with a main menu or new announcement.

The logout function is invoked when the student exits the system. During logout, data files containing student progress, completed work, event logs and email are uploaded or emailed to the course manager. Contextual information is added to the transferred files to assure the appropriate processing by the course manager.

5. Generating Web-Based Courseware Materials

ANDES is designed both to deliver distance courseware and facilitate the creation of such courseware. A special authoring language, called ATML, was developed to facilitate the generation of Web-based courseware.

Within an ANDES course, Web pages may be used to present new material, control interactive exercises, or deliver examinations. These pages should have a consistent form. For example, all pages may include a standard tool bar for navigating to other pages. Test pages might include a built-in timer widget, which locks the page after time runs out. A foreign language course might include multiple pages introducing vocabulary items, each of which contains standard elements like word definitions, examples of uses, and audio clips of spoken language. Consistency of form is achieved by generating pages from templates. The mini-server is configured for each course with a set of page templates appropriate for that course. Each defines a page layout, contains a set of fields to fill in, and defines the actions that the system should take when the student is done with the page. Each ATML page description identifies the template to be used, and lists the field entries. The field entries are indicated with tagged, SGML-like commands. The mini-server dynamically generates HTML pages from the descriptions, and inserts the appropriate CGI processing commands. Developers can combine ATML with HTML, or use just one or the other. ATML files can be created with any word processor.

As an example, consider a possible template for online examinations. Each exam includes a set of questions and a set of corresponding write-in text fields. The page has a title, a button to press when done, links to help pages, etc. Each exam has a fixed duration - exams submitted after the time expires will be discounted or rejected. Such complex behavior and layout can be generated automatically from the description shown in Fig. 4.

```
<@template>
<@type=open-exam>
<@title=Exam 1>
<@workshop=1A>
<@duration=60mins>
<@text=Exam on the first five lectures of
the course. You have one hour to complete this exam.>
<@questions=Describe the special effects that
can be used in film making./List and describe
the personnel associated with making a movie./Define
in your own words the term "key lighting.">
<@template>
```

Figure 4: Example ATML Template

6. Evaluation

Three phases of testing are planned for ANDES. In the alpha tests, described below, the emphasis was placed on the functionality of specific features of the system and the particular content included. In the beta tests, emphasis will be placed on the usability of the integrated user interface and the flow of student activities. In the first public course a wide variety of assessment techniques will be used, including analysis of the logs of student actions, correlation of patterns of student behavior with performance and satisfaction, time on task, errors, backward navigation, and so on.

6.1 Alpha Test Results

The alpha test was conducted March. At that time, the main student workstation functions and the login functions of the central server were complete. The evaluation lasted about twelve hours over two days and included a feedback session at the end of the last day. Four users, or "students", were stationed at two separate sites on campus and a teaching assistant was located at a third. The students watched the course video tapes and then logged in and took the workshops when appropriate. Eight workshops were tested. Each day concluded with an hour-long, remote TA session that utilized both Internet Relay Chat and Live Audio broadcasting.

When a student logged in, a connection to the central server was established and the student's workshop status was downloaded from the central database. The students were then presented with the main menu from which the workshops were launched; quitting the workshops took the students back to the main menu. At logout, another connection to the central server was made and student work was automatically emailed to the TA.

The feedback session after the evaluation focused on the content and efficacy of the courseware and systems technology. Specific problems and comments were recorded. The consensus was that the sum of the parts worked as a course, that the technology was easy to use, and that students would be interested in taking these "classes". Of particular interest were the chat sessions: The TA said that posing a question during chat was like asking a question in class and having all the students answer simultaneously; and that time would be needed to come up with a strategy for conducting these sessions. The students thought these sessions were effective but found their flow difficult to get used to at first. A maximum of ten students per session was suggested-only half the size we thought we could support. Most said the LiveAudio broadcast was interesting, but not particularly useful (in actuality it was used very little by the TA), although they liked being able to hear the TA's voice.

While the architectural functionality of the system was transparent to students during the evaluation, its setup was time-consuming and needs to be automated. Trying to configure the system to run in the required eight megabytes of RAM was especially troublesome. Ten megabytes are required just to support the Macintosh OS, Netscape, and Director, so virtual memory was used, which caused problems for some of the applications. The memory requirement will be increased to sixteen megabytes, which is fast becoming the multimedia standard.

References

[HTTP 96] HTTP Working Group (1996). Hypertext transfer protocol 1.0. Web page available at http://www. w3. org/ hypertext/WWW/ Protocols/HTP1.0/ draft-ietf-http-spec.html.

[Holt etal. 95] P. Holt, C. Fontaine, J. Gismondi, and D. Ramsden (1995). Collaborative learning using guided discovery on the internet. International Conference on Computers in Education, AACE. 206-213

[Johnson&Erdem 96] W.L. Johnson and A. Erdem (1996). Interactive explanation of software systems. Automated Software engineering. 3(4).

[Davis&Krafft 95] Building a production technical report server (1995). Advances in Digital Libraries, Spinger-Verlag, NY.

[Maurer&Lennon] H. Maurer and J. Lennon (1995). Digital libraries as learning and teaching support. International Conference on Computers in Education, 1995, AACE. 28-33

[Touch 96] J. Touch. Intelligent bandwidth (1996). Web page available at http://www.isi.edu/div7/ib/.

[Athabasca 96] Athabasca University (1996). Internet undergraduate courses. Web page available at http://www. athabascau.ca/ html/courses/online.htm.

[Open 96] The Open University (1996). Studying via the internet. Web page available at http://cszx.open.ac.uk/zx/.

Acknowledgements

The authors wish to thank the following individuals for their contributions to the ANDES project: John Zigrang of JTZ Engineering proposed the distributed server concept, and has contributed significant technical expertise and insight. Dennis Allard of Oceanpark Software assisted with the prototyping and database design, and advised on the PC implementation. Catherine Labore directed the development of the director courseware. Jay Krish, Jonghee Park and Behnam Salemi all contributed to the implementation of the system. This project is funded through a grant from the Annenberg Foundation.

Windows to the Universe

Dr. Roberta Johnson
Space Physics Research Laboratory
The University of Michigan, USA
rmjohnsn@umich.edu

Dr. Claudia Alexander, Cal Tech, USA,

Dr. Sandra Barlett, Turn of the Century Software, USA,

Mr. Michael Burek, Space Physics Research Laboratory, The University of Michigan, USA,

Mr. Theodore Clarke, Cal Tech, USA,

Dr. Joan Durrance, School of Information, The University of Michigan, USA,

Dr. James Green, Goddard Space Flight Center, NASA, USA,

Dr. Janet Kozyra, Space Physics Research Laboratory, The University of Michigan, USA,

Dr. Jon Linker, SAIC, USA,

Mr. David Mastie, Ann Arbor Public Schools, USA,

Mr. Paul Orselli, Long Island Children's Museum, USA,

Dr. Craig Rasmussen, Turn of the Century Software, USA,

Mr. Rivard Redding, Ann Arbor Hands On Museum, USA, and

Dr. Terry Weymouth, Turn of the Century Software, USA.

Abstract: Windows to the Universe is a World Wide Web site that presents information about the Earth and space sciences as well as related historical and cultural topics to the general public in an attractive and user-friendly way. The site makes extensive use graphically annotated button panels to allow intuitive navigation through the site. The site is graphics intensive, providing access to a rich archive of images, movies, animations, and data collected by satellites, spacecraft, and ground-based instruments. Intended primarily as an innovative information resource for museums, libraries, and classrooms, content within Windows to the Universe is developed to meet K-12 science education needs.

Introduction

Over the past 30 years, a wealth of remote sensing data has been obtained and stored by NASA and other Federal agencies pertaining to the Earth and space sciences. These data constitute a precious information resource, and are useful both for educational purposes as well as for development of value-added products with a wide range of uses. Results obtained from these data have generally been disseminated to the public through press releases and press conferences, for sale at museums, or in information booklets and posters which are available to educators upon request to the responsible NASA center or government agency.

More recently, with the growth of the National Information Infrastructure (NII), informational bulletin boards and home pages have appeared on computer networks which highlight various remote sensing databases and allow display of a variety of images and associated text. Interested users with connections to these networks and knowledge of the available tools and terminology can enjoy access to these resources, which are potentially useful and enriching to a wide-spectrum of the general public, particularly to students in the K-12 environment. Unfortunately, a large fraction of the general public does not have the computer literacy to navigate the information highway unaided in order to access these images. In addition, the user interfaces available, while elegant in their design as network tools, are not well suited to novice users. Many existing World Wide Web resources focusing on science are not specifically created in content and design for the novice user. These resources are usually written at the college level, using terminology not commonly in the lexicon of K-12 students. The important point is that a focused effort in the development of the presentation of content material and the crafting of the user interface is necessary before data on the Internet becomes information available for all users.

Although continually expanding in availability and use, Internet connectivity remains unavailable to a wide spectrum of the general public in their homes and schools. Museums and public libraries are prime locations for intersecting significant numbers of people, representing a broad cross-section of the general public ranging in age from pre-school to the elderly and including both sexes, multiple races, and diverse religions and nationalities. The yearly attendance at science museums nationally in 1993 was over 55 million persons. There are over 89,000 public and school libraries serving nearly every hamlet and community in the nation. These libraries are used by approximately half of the national population each year. These figures suggest that a focused effort on providing Internet access in libraries and museums may be a promising way to promote both technology and scientific literacy as well as to encourage recognition and use of the data bases available to the public. In recognition of the important role they have to play, libraries and museums throughout the country are increasingly including Internet connections and computer resources in their plans as budgets allow, and are highly interested in information resources that are appropriate for their settings. This development parallels the interest in and expansion of Internet connectivity and use in K-12 classrooms.

This paper describes the "Windows to the Universe" project, funded as a component of the NASA Public Use of Remote Sensing Data Bases Program (through the NASA Office of High Performance Computing and Communications Information Infrastructure Technology and Applications Program) to develop, implement, and deploy a network-based test-bed in libraries and hands-on science museums and designed to complement the curriculum needs of K-12 students of the Earth and space sciences. The goal of the project is to provide an attractive, easy and fun to use information resource in these settings as well as to attract and maintain the interest of casual, non-scientifically oriented users. After an initial description of the project, we discuss the content included in the site, the way the site is structured, educational uses, and plans for the future.

Description

Windows to the Universe (http://www.windows.umich.edu) is an innovative, user-friendly web site that guides users in museum, library, and classroom settings through informational text, images, movies, animations, and sounds to explore the rich archives of data collected by the satellites and spacecraft that have explored the Earth, the solar system, and the universe. Users can browse through the site (now including over 2000 individual html documents) in an intuitive fashion through the use of graphically enhanced button panels [Fig. 1] which illustrate the content available from a given page. The interest of the user is stimulated by elucidation of the historical and cultural ties between scientific knowledge and the human experience, as well as by providing information on new research discoveries, current events, and links to data bases of interest to users.

In order to be appealing to the broadest spectrum of the general public, primarily in informal learning settings, the site is highly graphically based. The over 1000 images and 80 animations and movies included within the site include hundreds of images of objects within the solar system (primarily from NASA), hundreds of images of astrophysical objects from NASA and observatories world-wide, historical photographs and movies from NASA missions going back to the beginning of the space age, photographs and paintings of famous scientists and philosophers dating back to ancient Greece and

Rome, photographs and artistic renderings of works of art from around the world relating to mythology about the planets, the stars, and conceptions of the universe, as well as numerous original images and animations created by artists working on the project.

Through their pre-college education, students are expected to develop an increasingly sophisticated understanding of science [American Association for the Advancement of Science 1993, National Research Council 1996]. Internet resources can make a significant contribution as information resources and potentially active partners to aid students as they develop these understandings through activities assigned by their teachers. We describe below in the section below on [Educational Use] how we are designing Windows to the Universe to meet the needs of the K-12 students and their teachers.

Figure 1: The main navigation panel for Windows to the Universe Version 2.0.

Windows to the Universe is being developed through a collaborative effort directed from the University of Michigan. Participants include the University of Michigan Space Physics Research Laboratory and the School of Information, Cal Tech, SAIC, the National Space Science Data Center (NASA/Goddard), Turn of the Century Software, the Ann Arbor Hands On Museum, and the Ann Arbor Public Schools. Our first prototype was released on 1 March 1996, and version 2.0 of Windows to the Universe was released on October 1, 1996. The application will continue to be expanded, with major revisions released every six months, to complete our planned content development by January, 1998. The following sections describe the topics and resources included in the site, how the site is structured, and how it is being designed to address pre-college science education needs.

Content

The disciplines spanned within Windows to the Universe include astrophysics and astronomy, solar and space physics, and planetary and lunar sciences, as well as the historical and cultural connections between this scientific information and the human experience. The left column of [Tab. 1] presents the main content areas that are accessible from the main navigation panel [Fig. 1]. For each of these main content areas, secondary navigation panel accessed through hyperlinks allow users to further specify their interests. Subdivisions of content provided through these secondary panels in the site are listed in the right column of [Tab. 1].

BEST COPY AVAILABLE

Historical and real-time images acquired using remote sensing and in-situ observation techniques from Earth-bound and orbiting telescopes, Earth and planetary satellites, and interplanetary spacecraft missions are utilized throughout the site. We also provide access to selected historical and real-time data and models of interest to users on topics such as weather on Earth and in space, earthquakes and volcanic activity, the Sun and the solar wind, and planetary and astrophysical observations. Animations based on readily available simulations of interesting time-varying phenomena such as the development of solar magnetic field structures, magnetospheric circulation and dynamics, and atmospheric structure are being developed and incorporated.

As is seen in [Tab. 1], we also include html documents and images about cultural and historical topics such as mythology, art and music, literature and films, and scientists or philosophers with connections to the Earth and space sciences. Examples include myths about gods and goddesses associated with the planets or constellations in different cultures, artistic renderings of myths or abstract conceptions of the universe, and music, literature, and films with themes relating to the Earth or space.

Structural Aspects

Navigation panels, consisting of a selection of graphically illustrated buttons corresponding to the different subject areas for which information is available, are used initially so that the user can intuitively select the area of interest to them. Version 2.0 of Windows to the Universe includes two levels of navigation panels prior to accessing content documents. Content documents typically include a descriptive title, a small image (which can usually be expanded through a link to a larger version of the same image), descriptive text with imbedded hyperlinks, a selection of links to other relevant resources within the site (text, images, movies, or sound), and a button bar for easy navigation to other main content areas within the site. In addition, we also present content in a "tour" format. The "tour" construct is used to allow a user to follow the development of a concept through to its conclusion, without providing hyperlinks which lead out of the tour, until the tour is completed. Options for movement on tours include forward, backward, back to the start of the tour, or return to the tour selection page.

The site is constructed to allow the user to specify the level of presentation they desire. In version 2.0, we provide three levels of content (beginning, intermediate, and advanced) which approximate the three levels of pre-college education [Educational Use]. The user specifies the level of presentation they wish as they enter the site, and this information is carried along and submitted to the server each time the user requests a document. Individual html documents within the site include all three levels of content, and the server uses the selected level (or "state") to process the document request and return the portion of the original document that corresponds to their level of interest.

To facilitate downloading of large images and animations, we have developed a CD supplement to Windows to the Universe Version 2.0 which will be provided at cost to users upon request. The CD includes most of the image, movie, animation, and sound files used within the site (but not the html documents), and will be updated on a regular basis as development on the site continues. This resource will then allow the client machine to download most of these large files directly from the local CD, rather than requiring transfer over the Internet. In this mode, the Internet is used only to transport evaluated html documents, images and data that update frequently (such as weather maps and images of the Sun), and images that our copyright agreements do not allow us to include on a CD.

Educational Use

Internet resources have great potential as a learning tool in appropriately equipped classrooms. [Tab. 2] summarizes concepts that students are expected to have mastered at various stages in their pre-college education related to the physical sciences, Earth and space sciences, and the history and nature of science according to the National Science Education Standards developed by the National Research Council [NRC 1996].

With this in mind, Windows to the Universe is being developed to meet science education needs of K-12 students and their teachers by: (1) developing content that covers many of the concepts outlined in [Tab 2], (2) developing content at three levels of sophistication meant to approximate the elementary, middle, and high school levels, including both use of appropriate terminology as well as graphics designed to attract students at these age levels, (3) developing content to supplement classroom units planned by classroom teachers participating in the project, and (4) developing a search tool for teachers that will allow them to specify grade level, concept or keyword, and the science standards document of their choice and have returned to them links to the pages within Windows to the Universe that are relevant and the pages of the science standards at which that concept or keyword are described. In order to insure that Windows to the Universe is a useful classroom tool, a team of six classroom teachers is participating in design of the site, selection of content, development and review of content, and evaluation of student and teacher use of the site.

Content Area	Content Subdivisions
Our Planet	Interior & Surface
	Atmosphere
	Magnetosphere
	Water
	Life
	Moons & Rings
	Planetary Characteristics
	News & Discovery
Our Solar System	Sun
	Planetary Systems
	Solar System Formation
	Interplanetary Space
	Comets
	Asteroids
	Solar System Characteristics
	News & Discovery
The Universe	Star Statistics
	The Big Bang
	Astrophysical Objects
	Stars
	The Warp Zone
	Stellar Evolution
	Sun
	Earth & Moon

Myths	Planets
	Sky
	Stars
	Infinity & Time
	Family Trees
	Cultures
Art, Music, Books & Movies	Space Art
	Space Music
	Space Books
	Space Movies
Space Missions	Space Exploration
	Manned Missions
	Unmanned Missions
	Future Missions
People	Ancient Epoch
	Middle Ages
	Renaissance Epoch
	Age of Enlightenment
	Modern Era
	Today's Scientists
	Timeline of Discovery
	News & Discovery
Data Sites	Space Science
	Planetary Science
	Astrophysics
	Earth Science
	Weather
Cool Stuff	Tours & Questions
	Hot off the Press!
	New Books & Movies

	Software & Games Neat Pictures
Headline Universe	Space Science News Astrophysics News Planets News Hot off the Press!

Table 1: Content subdivisions available from secondary navigation panels.

K-4	5-8	9-12
Physical science standards		
Properties of objects and materials	Properties and changes of properties in matter	Structure of atoms
Position and motion of objects	Motions and forces	Structure and properties of matter
Light, heat, electricity, and magnetism	Transfer of energy	Chemical reactions
		Motions and forces
		Conservation of energy and increase in disorder
		Interactions of energy and matter
Earth and space science standards		
Properties of earth materials	Structure of the Earth system	Energy in the Earth system
Objects in the sky	Earth's history	Geochemical cycles
Changes in Earth and sky	Earth in the solar system	Origin and evolution of the Earth system
		Origin and evolution of the universe
History and Nature of science standards		
Science as a human endeavor	Science as a human endeavor	Science as a human endeavor
	Nature of science	Nature of scientific knowledge
	History of science	Historical perspectives

Table 2: An overview of the physical science, Earth and space science, and history and nature of science content areas which pre-college students are expected to have mastered at various stages in their pre-college education [NRC 1996].

Plans

During the early phase of the project, two dedicated machines were deployed at the Ann Arbor Hands On Museum for evaluation purposes. We are now expanding participation in the project to other museums and libraries, as well as to K-12 classrooms. Interested teachers and representatives of libraries and museums with Internet connectivity and computers are invited to contact us to discuss participation in the project as the site continues to develop. User evaluations from focus groups in these libraries, museums, and classrooms as well as analysis of our on-line comment forms input and server logs will continue to be used to guide revisions to the current and future revisions of the site.

References

[American Association for the Advancement of Science 1993] Benchmarks for Science Literacy (1993). Association for the Advancement of Science, Oxford University Press, Inc.

[National Research Council 1996] National Research Council (1996). National Science Education Standards. National Academy Press, Washington DC.

Acknowledgments

This project is funded by NASA grant NAG 502876.

A Browsing System Based on Multimedia Cohesion

Jim Jungbauer
Department of Computer Science, Campus Box 430
University of Colorado at Boulder
Boulder, CO 80309-0430 U.S.A.
(303) 492-8136; jungbaue@cs.colorado.edu

Patricia Baggett
Department of Mathematical Sciences
Science Hall Room 236, New Mexico State University
Las Cruces, NM 88003-8001 U. S. A.
(505) 646-2039; baggett@nmsu.edu

Andrzej Ehrenfeucht
Department of Computer Science, Campus Box 430
University of Colorado at Boulder
Boulder, CO 80309-0430 U. S. A.
(303) 492-7212; andrzej@cs.colorado.edu

Abstract: A system for browsing multimedia information is presented. The data to be browsed are organized according to cohesive elements, features that are common to a group of data. The system employs rules that force browsing along cohesive element paths; that is, through pieces of data that share a cohesive element. The complete system consists of two software components, an authoring tool and a browser. The authoring tool is used to create a presentation of information according to organization rules based on cohesion. The browser is then used to traverse the presentation created by the author. We are interested in the effects of organization of multimedia information by cohesive elements on the ability of the user to learn about the material being presented. Other research questions and the current state of the implementation are discussed.

I. Introduction

What is a browser?

We talk about browsing when the user can recognize valuable or relevant information but is unable or unwilling to ask for it. Cohesion (or coherence) is a term that originated in psycholinguistics. Cohesive elements of a text are key ideas that are repeated in the text over and over again, and which form a main topic of discourse. High cohesion distinguishes a meaningful text from a random collection of sentences. In general, a common feature of a group of items may be called a cohesive element for the group.

Examples of browsing in everyday life

1. An old book collector's dream is to find a valuable book that is not listed in any catalog, but which is virtually unknown. Of course one cannot ask for such a book, but one may find it by browsing in an antique store between cookbooks and old photos, on a shelf labeled miscellaneous.

2. Keys to flowers and trees are not generally organized according to scientific classifications. The structure of a flower, which is the basis for scientific classification, requires expert knowledge, and it is often hard to observe. On the other hand, an experienced amateur can recognize that a specific picture and description of a species match a given specimen. A typical flower guide groups flowers by colors. A color, such as lavender, is a cohesive element for each group. Within each group, the flowers are arranged by placing the most similar ones next to each other. The user may choose the color on the basis

of the specimen at hand, and browse through the list until he or she finds the right match. There are other possibilities for organizing a key. Cohesive elements may be geographic regions, type of habitat, season of flowering, and so on.

3. Very few supermarket shoppers have a shopping list that specifies each item completely. Most shoppers go to the proper aisle, e.g., dairy products, canned vegetables, toiletries, and browse to get the best bargain. (Stores encourage browsing because browsers buy more.) Here, the cohesive elements for each aisle are usually clearly marked and displayed, such as bakery goods, Mexican foods, and so on.

The common theme among the above examples is that items are grouped by cohesive elements, and users browse by scanning the items within one group. Users do not choose the cohesive elements; instead, the elements are selected by store owners, supermarket managers, or authors of flower keys. For example, in a supermarket one can put together canned goods, mixing meat products with vegetables, or one can put together vegetables, mixing together fresh and canned produce.

Learning by browsing along a cohesion path

There are some central ideas which form a cohesion path that crosses different disciplines. Preservation of energy is a central idea of physics, chemistry, cosmology, and many other domains. Organizing learning along such ideas counters fragmentation and over specialization, which are characteristic of modern scientific development.

Besides classifying information by content, we can classify it by media: English text, Spanish text, audio, movie, graphic, video, interactive video, and their combination. Media may also act as cohesive elements which correspond to grouping information by form of presentation. Different media are often difficult to combine. In a book about birds, we have text and pictures, but we cannot include chirping or acrobatic flying, which can be heard and seen in a movie.

In all sequential media, e.g., written or spoken text, movie or video, each item can lie on only one cohesive path available to a browser. If flowers of one color are put together, then flowers from the same region must be scattered throughout the book, and vice versa. Using hypermedia to construct a browsing system is a very attractive solution to this dilemma. Then the user of a system may choose the browsing path, following color (blue), or blooming season (April), or habitat (desert), getting the advantage of all cohesive elements within one tool. Also existing hypermedia techniques can overcome the boundaries among media, allowing their simultaneous combination.

The goal of this project is to apply the concept of cohesion to a multimedia browsing system. The system we are developing is described in detail in III. below.

II. Previous research

A. Psychological studies

As mentioned above, the concepts of text coherence and cohesion have been used by many authors in the psycholinguistics literature ([Halliday 1985]; [Halliday & Hasan 1976]; [Grimes 1975]; [Harris 1952]; [Kintsch 1974]; [Kintsch & vanDijk 1978]). Two items in a text are said to be cohesive if they refer to the same concept. In the schema of [Halliday & Hasan 1976], examples of cohesion include repeated words, a noun and its pronoun referent, and the use of synonyms. These authors claim that cohesion provides semantic continuity and improves a text's intelligibility. An extended treatment of text coherence can be found in [Kintsch 1974].

The notion of text cohesion was extended to visual, and to visual and verbal cohesion by [Baggett & Ehrenfeucht 1982]. In their schema, two items from different media (textual and visual) were termed cohesive if they referred to the same concept (e.g., the word 'husband' and a picture of a 'husband'.) They described a cohesion graph for frames from a movie; with it, a measure of the closeness of cohesive elements in a sequence of frames could be determined. In one experiment, participants were asked to put still frames from a movie that they had not seen into a sequence so that they "make a good story".

Participants grouped together cohesive elements significantly more closely than they actually occurred in the movie. The authors interpreted this result to mean that people are aware of cohesive elements and use them in grouping material.

B. Some earlier implementations

Here we briefly describe some of our previous implementations of browsing systems based on multimedia cohesion. (Discussions of other approaches to multimedia browsing can be found in [Allen 1995]; [Cruz & Hill 1994]; [Stotts & Furuta 1989]; [Zellweger 1989].)

(1) [Baggett, Ehrenfeucht, & Guzdial 1989] implemented a graphics-based instructional system for assembly and repair of a toy "string crawler", a battery-powered vehicle which traveled along a string when it was turned on. Access was via graphical and textual cohesive elements. Users could follow any of 21 different concepts (e.g., switch box, wire, motor) through the 41-frame presentation.

(2) In "Where in History is Mathematics?", a project in two graduate software design classes, users could investigate the six different civilizations which, according to current knowledge, developed mathematics at least in part independently: Africa, Sumer-Babylon, India, China, Maya, and Inca. Cohesive elements included zero, counting schemes, famous individuals, games, and tools.

(3) "Footloose in Washtenaw County, Michigan", another project in a software design class, was a multimedia (videodisc-based) implementation of sites one could visit. Cohesive elements included cross-country ski areas, bird-watching locations, and old churches.

(4) In an implementation of 183 Emily Dickinson poems and critiques of them offered by four authors [Espinosa & Baggett 1994], cohesive elements were 16 abstract themes (e.g., love imagery) together with the poems themselves. The material in this system was text only.

It should be noted that none of the above implementations was a complete browsing system. They did not provide an authoring tool, so each one was restricted to being used as a browser of a specific presentation. The current system being developed is much more general, as described below.

III. Current design and implementation

A. Users of the system

In all browsing schemas we have two participants: an author, an expert who arranges the material and identifies or creates cohesive paths, and a learner who chooses and follows a path in search of information. In an experimental setting, there is a third participant, a researcher, who can observe the learner (or the author), analyze their behavior and the workings of the system, and plan experiments.

B. Data structure

The material to be browsed will be organized into a data structure, essentially a graph, that groups cohesive materials together. Each multimedia document is assigned to a node in the graph (possibly more than one document per node), and nodes that share a cohesive element are linked together into a cohesive path. Nodes may belong to more than one cohesive path, and one special node, the home node, belongs to every cohesive path. Each path is circular and has a default order on its nodes so that the home node is both the first and last node in every cohesive path. Since nodes may belong to more than one cohesive path, these paths may intersect. The default ordering of nodes within a cohesive path will determine the order of forward traversal when the graph is used to browse the material, and the intersection points (nodes) are where travel along one path may yield to travel along a different path.

C. System components

This section describes the current implementation of our browsing system. Three components have actually been designed, each of which corresponds to one of the three types of users described earlier.

The first to be implemented is the authoring tool, which is used by the author to create the presentation to be browsed. The second piece being developed is the browser, with which the learner will browse the presentation. The final component consists of facilities for the researcher to analyze the learner's session in the browser. At the present time, only the authoring tool has been partially implemented.

1. Authoring tool .

This tool provides the functionality needed for the author to create a multimedia presentation. The underlying data structure for a presentation is the cohesion-based graph discussed previously. Because of this, the terms "graph" and "presentation" will often be used interchangeably here.

To facilitate the author's creation of a good presentation, the authoring tool provides the author with services for both constructing the underlying graph and analyzing the graph that has been built so far. With respect to graph construction, the author has three primary tasks:

(1) assign the materials to be browsed to individual nodes;

(2) decide on cohesive elements among the nodes; and

(3) establish the order of nodes within each cohesive path.

The authoring tool accommodates the author by providing this functionality. In order to assign the material to nodes, several node-level operations are available. The author may create new (initially empty) nodes and delete existing nodes. The author may also edit an existing node by editing its default text or its document references, which point to any multimedia documents associated with the node. To establish cohesive relationships, the author may add and remove nodes to and from cohesive paths corresponding to specific topics. (Because cohesive paths correspond to lists of nodes sharing a common topic in the presentation, they will often be referred to by other interchangeable names, such as "cohesive topic lists" or "topic lists." Similarly, we use the terms "topic" and "cohesive element" interchangeably.) The author may determine the order of nodes within cohesive paths by viewing the topic lists and rearranging the nodes in these lists. The author may also manipulate the graph at the topic level. The author may create and delete topics, as well as add and remove entire topic paths to and from the home node of the graph. This effectively allows the author to include or exclude entire cohesive element paths. In this way, the author may experiment with changing the presentation at the path level while still keeping unused paths around to be "attached" when needed.

As for analyzing the current graph, the author may perform several operations on the graph to determine some of its properties, such as the longest cohesive path, the largest number of links to any node, or the longest or shortest path between two nodes. Some automatic checking of the current graph will also be done by the authoring tool. While constructing the presentation, all operations (e.g., adding and removing nodes and topics) will be checked by the system to ensure that the author may not manipulate the graph in such a way as to make it inconsistent with our model of a legal cohesion-based graph.

2. Browser.

The product of the authoring process is a completed presentation. This presentation is input to the browser, which allows the learner to traverse the graph and browse the materials it contains.

Traversal. The browser allows the learner to traverse the graph in a rather restricted way. The basic rules of traversal are as follows. A learner may:

(1) move to the next node within the current active path;

(2) move to the previous node within the current active path;

(3) move to the next node in an intersecting path (when currently at a node belonging to multiple cohesive

paths);

(4) move immediately to the home node; or

(5) backtrack through the most recently visited nodes.

The notion of the "active path" refers to the path corresponding to the most recently selected cohesive element. When moving to the "next" or "previous" node from a node belonging to more than one path, the learner will move to the node belonging to the same topic that had been followed up to that point. Aside from jumps to home and backtracking, the learner's traversal may be summarized as follows: a learner must move one node at a time and must remain on the same cohesive path, except when at a path intersection; at an intersection, a move to the next node on any intersecting path is allowed. Note that the learner has the choice of any topic path in the presentation when at the home node. A good analogy for this type of traversal is travel along a bus route. A rider may continue on a given bus line or may transfer to another route at any stop where routes intersect. A possible advantage to this method of traversal is that the learner is not forced to make decisions about what topic to select next (except when starting at the home node). He or she may simply follow the default ("next") links through a cohesive path, although the option of leaving the path is available at intersection points. This reduction in decision making may allow for more focus on the material rather than on the traversal.

Node display.

In addition to enforcing the rules of traversal, the browser is of course responsible for displaying the documents associated with each node. It is left to the author to determine what materials are included with each node, but in any node it is possible to have default text that is automatically displayed when the node is visited, as well as supplemental links to documents of many other media. These are referred to as supplemental links to distinguish them from traversal links that connect nodes. Supplemental links may be used only to access multimedia documents (e.g., play a sound file or a video clip) associated with a given node, and not to move from that node to another part of the graph. They can be thought of as extending the node. Each node will generally correspond to a one page on-screen display. Upon arriving at a node, the learner will typically see the default text for the node, the mandatory traversal buttons (Next, Previous, Home, and Back), any intersecting path transfer links (if at a node belonging to multiple cohesive paths), and any supplemental document links. Each supplemental document may have its own display page as well, which will of course provide a mechanism for returning to the default page for the node.

3. Research Facilities.

The primary mechanism available for examining a learner's use of the browser during a session will be a log file in which all learner actions will be logged and time stamped. This record will allow analysis of a learner's behavior, such as the amount of time spent on certain parts of the graph and which nodes were accessed most frequently and in what order. In addition to the log file, it is possible that direct observation of browsing sessions will be used in the analysis, as well as post-session interviews with the learner.

IV. Experimental parameters

Some of the characteristics of the system we are developing cannot be determined without first using the system in an experimental fashion to help determine some parameters. This is particularly true of the browser, whose features and layout may significantly affect the behavior of the learner. There are also parameters associated with the actual presentations being designed that will have to be varied in order to determine what is reasonable and usable for the learner. Some examples of such parameters to be determined experimentally are:

* The number of cohesive elements per node

* The distribution of cohesive elements over nodes:

Should some nodes act as hubs belonging to many cohesive paths?

Or should there be an even distribution of topics per node?

* The distribution of supplemental documents over nodes:

Should many supplemental multimedia documents be linked to one node?

Or should each multimedia document be on its own node?

V. Research questions

We are also interested in broader questions about the value of a browsing system based on cohesion for both learners and authors. For example,

What are the benefits of this kind of active exploration for the learner?

Will some cohesive paths be followed more frequently than others?

Will the learner get lost?

Will the learner find what he or she is seeking quickly and efficiently?

Will the author be able to influence the user's learning by the manner in which the cohesive elements are selected and linked?

References

[Allen 1995] Allen, R. (1995). Interface Issues for Interactive Multimedia Documents. Forum for Research and Technology in Digital Libraries , 1-12.

[Baggett & Ehrenfeucht 1982] Baggett, P. & Ehrenfeucht, A. (1982). Information in content equivalent movie and text stories. Discourse Processes, vol. 5, 73-99.

[Baggett, Ehrenfeucht, & Guzdial 1989] Baggett, P., Ehrenfeucht, A., & Guzdial, M. (1989). Sequencing and access in interactive graphics-based procedural instructions. University of Michigan School of Education Technical Report for Office of Naval Research, vol. 2 (1), August.

[Cruz & Hill 1994] Cruz, G. & Hill, R. (1994). Capturing and playing multimedia events with STREAMS. Proceedings ACM Multimedia, 193-200.

[Espinosa & Baggett 1994] Espinosa, R. & Baggett, P. (1994). A design for accessing multimedia information using cohesion. In J.D. Brouwer-Janse & T.L. Harrington (Eds.), Human-Machine Communication for Educational System Design, Eindhoven, The Netherlands: NATO Advanced Study Institute, Springer-Verlag: Series F: Computer and Systems Sciences, Vol. 129, 275-282.

[Grimes 1975] Grimes, J. (1975). The thread of discourse. The Hague: Mouton Press, 1975.

[Halliday 1985] Halliday, M. (1985). An introduction to functional grammar. Baltimore, MD: Edward Arnold.

[Halliday & Hasan 1976] Halliday, M. & Hasan, R. (1976). Cohesion in English. London: Longman Press.

[Harris 1950] Harris, Z. (1950). Discourse analysis. Language, 28, 1-30.

[Kintsch 1974] Kintsch, W. (1974). The representation of meaning in memory. Hillsdale, NJ: Earlbaum.

[Kintsch & vanDijk 1978] Kintsch, W. & vanDijk, T. (1978). Toward a model of text comprehension and production. Psychological Review 85, 363-394.

[Stotts & Furuta 1989] Stotts, P. & Furuta, R. (1989). Petri-net-based hypertext: Document structure with browsing semantics. ACM Transactions on Information Systems, v. 7, 3-29.

[Zellweger 1989] Zellweger, P. (1989). Scripted documents: A hypermedia path mechanism. Proceedings ACM Hypertext, 1-14.

Acknowledgment

We thank the Office of Naval Research for funding this work under grant number N00014-95-1-0858.

On-demand Multi-lingual Font Service on Heterogeneous Computer Platforms

Wing Kay Kan (wkkan@cs.cuhk.edu.hk)
Stephen S M Wong (smwong@cs.cuhk.edu.hk)
Vincent S Y Yiu (syyiu@cs.cuhk.edu.hk)
Gilbert H F Young (young@cs.cuhk.edu.hk)
High Performance Computing Lab
Department of Computer Science & Engineering
The Chinese University of Hong Kong
Hong Kong

Hak Wai Chan (hwchan@cs.hku.hk)
Henry W H Cheung (whcheung@cs.hku.hk)
Department of Computer Science
The Hong Kong University
Hong Kong

Abstract: The plug-in modules described in this paper supports World Wide Web (WWW) browsing in multiple languages, such as Chinese, Japanese, and Korean language. This project aimed at creating software tools for WWW browsing that does not limited by the availability of local fonts installed on client computers. As most computers sold to end users nowadays are come with English font set only, very few are equipped with other fonts in local languages. The problem to display non-English languages through WWW service will be discussed. A novel solution to solve the font problem will be proposed. With the product in this project, WWW browsers can display multi-lingual documents independent of the character and font sets on heterogeneous computer platforms.

Background

The Internet has been a new paradise for information exchange. In the cyberspace there is no nationality barrier. There are more than 100 countries connected to the Internet and tens of millions users are using it. People from different countries are free to distribute and receive information through different channels such as newsgroups, gopher, archie and emails on the Internet. Among these channels, WWW supports graphical user interface, it is easy to learn and easy to use. It also supports multimedia contents.

The existence of WWW and the Internet supports communication and information exchange breaking the geographical limitation among people in the world. However, inconvenience has been observed in non-English speaking places when local characters are required in WWW information pages. To display a (foreign) language, the character and font set for that language had to be installed on a local font server or on the local machine. English is the most popular language in computer industry such that every computer is expected to be able to display. However, it is not the case for other languages. For example, it is unreasonable to expect the majority of computers designed for Greek speaking area to have Korean font/character set be installed.

There are a few methods to solve the multi-lingual display problem. One popular solution is to store and transmit graphical images instead of characters. Every page in a WWW server is stored in a graphical format, such as JPEG or GIF format. The (foreign) characters become dots and pixels on the graph. However, this method degrades the speed of WWW browsers by many times, because it takes many bytes to represent the same character in pixel form.

Another method to solve the multi-lingual display problem is to install fonts on client computers. It is one of the most popular methods in use for WWW documents with local characters. There are quite a few problems with this approach. Firstly, not all users have the required expertise to install fonts properly onto their computers. Secondly, font files occupy considerable hard disk space. Thirdly, some client computers may be owned by an organization, and hence managed by dedicated personnel. End users may not have privileges to install font files onto these computers.

Some of the existing WWW browsers, for example, the Netscape Navigator Version 2.0, are able to display a number of language encodings like Chinese (Big 5, GB), Japanese (EUC-JP, SJIS) and Korean (ISO-2022-KR, EUC-KR). However, there are two limitations for them.

(1) They are not able to display mixed-language documents, that is, a single document consists of more than one languages such as Chinese, Japanese and Korean will not be displayed properly. The selection of language encoding is an explicit task to be done by end user.

(2) If the client computer or the local font server does not have the font installed, the browser will not be able to display document using that particular character/font set.

On-demand Font Service

In order to solve the multi-lingual document display problem, an on-demand font service that does not require system set-up and configuration is proposed. The on-demand font service should have the transmission speed no worse than the graphical method as stated in the Background section. The font service should be able to handle multiple font sets in different languages. In summary, the design goals of the proposed service are the following:

(1) Requires no privileged system set up

(2) Capable to handle multiple languages

(3) Acceptable speed in operation

The first criterion is handled by using standard WWW browser supporting Java applets. There are a few Java-capable WWW browsers on the market. The merits of using Java applets will be discussed in [Merits of Using Java Applet].

The second criterion is to allow multiple languages to be displayed. This goal is partly achieved through document syntax used on the server side, and another part is the Java applet developed to support different character and font sets with the same internal codes. That applet will also be able to display characters in different font families, size, and attributes.

The third criterion is handled partly by the 'on-demand' nature of the font service. Unlike the method to install foreign font files on client computers, which font image for all characters in a particular language is downloaded and installed, only the font images of the characters actually used in the WWW document are transmitted on the network. The improvement in speed, compare with the font file installation method, will be significant for lots of Asian languages, especially for Chinese, which are composed of tens of thousands of different characters.

The speed is further improved as there are expected to have repeated characters in a document. In the proposed on-demand font service, for the same character, the font image will only be transmitted once, and all further occurrence of the same character will be displayed by using the font image already transmitted and cached into client computer memory. This improves the speed on WWW browser compare with document in graphical format.

Merits of Using Java Applet

Java [Hoff et al. 96] is an Object Oriented (OO) language developed by Sun Microsystems for distributed computing environment.

One of the advantages by using Java is the machine neutral property of the compiled byte-code. That means, the same applet can be used on heterogeneous hardware platforms. The compiled byte-code is efficient enough in speed to be comparable to machine code compiled by native compiler.

Java supports multi-threading in the language syntax. Multi-threading is useful in this project, as we can partition the functions performed by Java applets into several sub-tasks. One of the thread deals with the network in order to load the required fonts on demand. Another thread processes the document requested to be displayed on screen. Yet another thread looks for user interaction and responds accordingly.

Security is an integral part in Java applet design. There are several ways in a Java applet to safe guard crackers, Trojan Horse or computer virus programs. Java applet is normally not allowed to make network connection but only to the host where the applet is come from. This measure can stop Trojan Horse programs that intend to gather local information. Java applet is not allowed to read/write files on local computer, for the same reason to stop Trojan Horse. Java is a strong-typed language. All references to memory should go through a defined object, so that the size and limit of an object can be verified.

The On-demand Font Service in this project is implemented by using Java applets. We believe that Java provides a suitable environment for us to develop a distributed network tool which can be run on heterogeneous computer platforms. The Object Oriented features in Java language helped us to write manageable code. The security features in Java applet assures the safe operation in a public network. The multi-threading ability in Java language let us exploit available computing resources without paying extra effort to cater for different computer architectures. Java is new and hot, we foresee that the Java language may be one of the popular languages in distributed computing environment in the future [Ritchey 95].

System Setup

A complete On-demand Font Service on Heterogeneous Computer Platforms composed of a number of entities. They are:

WWW Server

The WWW server is responsible for store and transmission of WWW documents. On the WWW server, it runs a daemon program, usually called `httpd`, which it listens on a network socket for HTTP request. The syntax and response for the request are defined in the Hypertext Transfer Protocol [Lee et al. 96]. In the WWW server filesystem, it stores multi-lingual documents, the on-demand font service Java applets and the characters/fonts are supported by that server. There is no special setting nor non-standard programs on the WWW server for this project. The WWW server even does not require to support multiple languages in its own environment.

Multi-lingual Document Syntax

In a WWW page, if there are multi-lingual documents in it, the WWW administrator has to specify them by using a defined syntax. As existing WWW browsers support Hyper Text Markup Language (HTML) [Lee et al. 96], the syntax to define a multi-lingual document follows that standard. The following is a sample coding to include a multi-lingual document called `lang_file` containing traditional Chinese (BIG-5 encoding) characters:

```
<APPLET CODEBASE="java" CODE="FontServer.class" WIDTH=600 HEIGHT=1000>

<PARAM NAME=docname VALUE="lang_file">

<PARAM NAME=encoding VALUE="big5.hku">

<PARAM NAME=fontsize VALUE="12">
```

```
<PARAM NAME=typeface VALUE="Song-normal"> </APPLET>
```

The applet `CODE` parameter specifies that an applet called `FontServer.class` will be transferred from the WWW server to the browser. The parameter `docname` specifies the multi-lingual file to be displayed. It can be a plain text file using character `encoding`, `fontsize` and `typeface` as indicated by the corresponding parameters.

The applet `FontServer.class`, upon received the above parameters, will be run on the client computer, invoked by the browser. The applet will request the multi-lingual file (in the above example, `lang_file`) from the WWW server, and start to display that file on the client screen. If there is no font on the client system matching the encoding of the multi-lingual file, the on-demand font server will load the images of required characters from the WWW server.

The multi-lingual file should be a plain text file as supported in the first version of the On-demand Font Service. Multi-lingual file using more than one languages has to be declared by using more than one of the above applet declarations. See [Fig. 1] for a sample screen by using the On-demand Font Service Applet.

In later versions, a subset of HTML commands will be supported in the multi-lingual file, which the goal is to let the document switches language encoding, font size and typeface on the fly. The parameters coded in the applet declaration will then be the default values for the multi-lingual file.

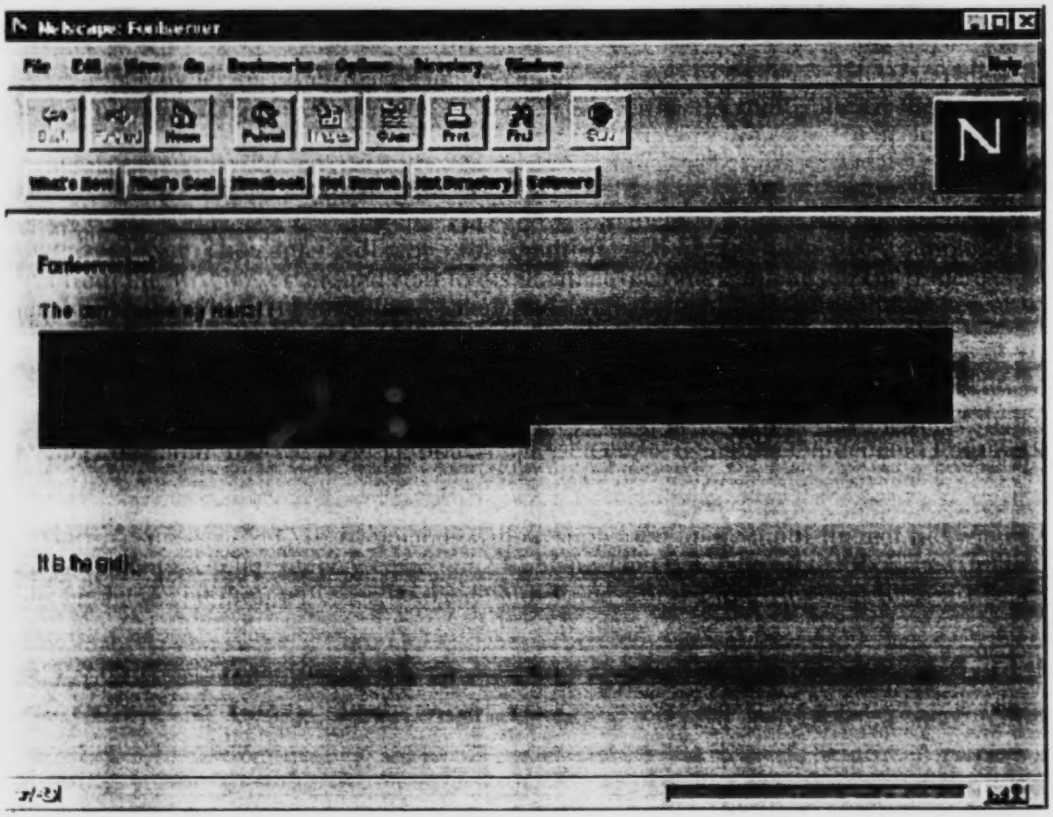

Netscape: Fontserver
File
Edit
View
Go
Bookmarks
Options
Directory
Window
Help
Back
Forward
Home
Reload
Images
Open
Print
Find
Stop
What's New
What's Cool
Handbook
Net Search
Net Directory
Software
N

Figure 1: Sample Screen by Using the On-demand Font Service Applet

Multi-lingual Font Location

BEST COPY AVAILABLE

As the `FontServer.class` applet will on-demand load font images, the WWW server should have files to support all languages used by multi-lingual documents. The following is the directory structure to store fonts:

`fonts/<encoding>/<typeface-size>`

Here is a sample directory tree:

`fonts/big5.hku /song-12/`

`/song-20/`

`/gb2312.1980/kai-12/`

`/kai-16/`

The `fonts` directory can be an alias defined in the WWW server. Inside each sub-directory, there are font image files using the internal code of the character as file names. However, the file name convention is considered to be proprietary to the on-demand font server; as in later versions, the arrangement of font image file names may be changed.

WWW Client Set Up

On WWW clients, a WWW browser is invoked by the end user to view multi-lingual documents. As stated in the above sections, the only requirement of a suitable WWW browser is that the browser should be Java-capable. Netscape Navigator Version 2.0 and Hotjava are examples of Java-capable WWW browser.

Font Service Applet

A Java-capable WWW browser, when encounters the `FontServer.class` applet tag, will request the On-demand Font Service applet from the WWW server. After receiving the `FontServer.class` applet, it will be started by the WWW browser. After the `FontServer.class` applet takes over the control, it will examine the `docname` parameter, and the retrieval of the multi-lingual document will be started.

When the `FontServer.class` applet receives the multi-lingual document, it will analyze the character/font requirement of the document and check if the local site has the character/font installed.

If local site has the character/font set installed, the `FontServer.class` applet will use it. Otherwise, the `FontServer.class` applet will further request the character/font files from the WWW server. Note that only the characters used in the multi-lingual document will be transferred.

Future Developments

The on-demand font server is just the first stepping stone to a complete versatile multi-lingual WWW environment, we have plans to enhance the system in the following areas.

Input method support is an important part in a truly multi-lingual environment. Our on-demand font server solves the display problem in non-English based information documents. From time to time, users may want to type feedbacks to the WWW server, or to search for a particular word in the retrieved document. The way to support input methods for non-alphabetic languages is considerable more complex than simply read the codes generated from the keyboard. We will study the possible mechanisms for On-demand Input Methods in future papers.

There is inherited redundancy in almost all human languages, if we can exploit that property, then we can devise a *good* font load order in order to let users to understand a document page even before all the character fonts are loaded. We already have a few ideas which show promising results. We will implement an intelligent font load order in future version on our on-demand font server.

We believe that by compressing the raw font images, we can achieve a faster load time for on-demand font retrieval. Another team in our lab which is specialized on compression technique will explore possible ways to compress font files to a great extent.

In the first version, the multi-lingual file is a plain text file, we aim at a better support of HTML syntax in later versions. We would like to see the multi-lingual file can support more than one page of text, on-the-fly language encoding switching, font size and typeface changing, and eventually, to support hyper-link in the multi-lingual document.

In the first version, all multi-lingual fonts are in bit map format, which requires the minimal support on the client side to display properly on screen. However, vector font will let users have a better control on the font size and appearance. We will explore how to make use of vector fonts on our on-demand font server to get acceptable display speed.

Conclusion

Our team has proposed a language independent, on-demand font server on heterogeneous computer platforms. The significance of the project is to let computers connected on the Internet around the world to be able to display any foreign language WWW documents efficiently. Our implementation makes use of the Java language to add intelligence on the client side and manage the on-demand loading of fonts from a WWW server. Our design does not depend on configuration modifications on the client side and the software running on the server side is standard available version. It is important as our software can be used by non-privileged users running on-the-market software. We have proposed a few areas that this project can continue to make a truly multi-lingual environment.

References

[Hoff et al. 96] Arthur van Hoff, Sami Shaio, Orca Starbuck. Hooked on Java. Addison Wesley 1996.

[Lee et al. 96] T. Berners-Lee, R. Fielding, H. Frystyk. Hypertext Transfer Protocol -- HTTP/1.0 Draft 5. INTERNET-DRAFT 1996.

[Ritchey 95] Tim Ritchey. Java!. New Riders Publishing, Indiana 1995.

Acknowledgments

This project is partially supported by the Hong Kong UPGC RGC Grant number 338.065.0016.

Real Audiences-Worldwide: A Case Study of the Impact of WWW Publication on a Child Writer's Development

Judy Keiner
Senior Lecturer in Education
Department of Education Studies & Management
University of Reading
Bulmershe Court
Woodlands Avenue
Reading
RG6 1HY
England
J.C.Keiner@reading.ac.uk

For many years, theoretical and practical work on the development of children's writing has stressed the importance of devising tasks which require children to write for real audiences for real purposes, as opposed to textbook exercises completed for teachers. In the world of paper-based communication, this has emphasized the importance of publishing as a concept. Publishing is taken to include the processes of drafting and redrafting for the needs of the identified audience, usually involving the creation of an artefact in some way resembling a commercially produced book, often with a hard cover. Typical examples have been the production of a story by an older child for a younger audience, or a story or information book intended for display or sale to a school-based, parental or wider community audience. At a theoretical level, the key features of publication by child writers for real audiences are, firstly, a shift from a one-to-one mode of communication to the one-to-many mode typical of real-world publication. Secondly, there is a strong focus on emulating the presentational features of commercial or other adult publication, notably the bound book format. The advent of laser-printed desktop publishing and inexpensive colour copying facilities has added to the opportunities for child authors to emulate commercially produced "real books", though there has been relatively little discussion of this in the literature.

Sites now exist on the WWW which specialize in the electronic publication of writings, art and photographic work by children. Many hundreds of children worldwide are having their written work published on them. Little has so far been published on the impact of these facilities on the child writers who use them. Most work so far published on computers and children's writing has focused on computer-originated communication, particularly the opportunities and constraints offered by word-processing and desktop publication.

The opening up of e-mail facilities for use by children has also been the subject of some analytic work. For the most part, the documented use of e-mail by children has been in the context of school-based projects. These have typically involved one-to-one or many-to-one dialoguing exchanges. Examples have been projects involving classes of children exchanging information about a shared topic, or a single class consulting an expert, or sometimes a simulated external event to which the children respond, such as messages apparently arriving on their screens from visitors from other planets .

The advent of WWW as a medium for children's publication opens up a venue for real world publication of a new order. WWW publication most readily replicates for child authors both the one-to-many locus of the adult published author, and in addition allows for the one-to-one medium of individual mail responses, given the facilities offered by WWW sites for children of hyperlinked facilities for direct email response by readers.

This discussion focuses on a case study of one writer, S, a girl aged between nine and ten over the period of six months analysed. The developments analysed in this paper took place in her school and mainly in her home. S's school had no online facilities on its computers for pupil use, and only crude word-processing facilities with dot matrix output, which were in any case rarely used.

In the spring of 1995 her mother acquired for the first time a multimedia computer with facilities to access WWW and email. Acting as mentor for her daughter's induction into computer mediated communication, she sought to interest her daughter in the newly acquired medium. Initially a talented but very reluctant writer, and a very advanced and voracious reader of books, S initially showed little interest in WWW, although she was shown a number of sites related to children's interests over a period of several months. She began to show interest in a number of school and dedicated sites publishing children's writing and artwork after a period of some three months of occasionally being shown various sites selected by her mother, and having opportunities to follow links to these and other sites. At no time did she show any interest in email, although the possibility of acquiring "keypals", analogous to penpals, was offered to her.

One of the sites in which she showed a particular interest was the US-based KidPub site. This site accepts all stories submitted by or on behalf of children and young people under the age of 16 and publishes them, each text being given its own Web page. Each text carries the title and name of the author, and, where submitted, a photograph of the author. At the end of the text appear any details submitted about the authors, such as age, place of residence, hobbies, etc. Where available, a link to the email address of the author is provided. Finally, at the end of the Web page, the number of times the story has been accessed is given, in the format "This story has been read n times". Visitors accessing the KidPub site are offered lists of all texts available, divided into "Newest stories" and stories previously submitted. These are organized into roughly monthly volumes. The individual stories are listed as a series of links giving only title, name and age of author and place of residence, usually by hometown and state or country. Further choices offered include a section devoted to stories submitted by schools, as well as opportunities to submit stories direct via email response links. KidPub is a popular site, receiving some hundreds of accesses per day. It is also featured as a link on a number of compendium sites listing links of interest to children.

S showed interest in visiting the site, but was not responsive to being invited to compose a story specifically for it, or initially to having one of her existing stories transcribed and submitted. Her interests when browsing the site were invariably to look at the current day's stories, rather than at previous volumes. She was intrigued by the counter facilities, showing interest in how many times a story had been read. She tended to choose to read stories submitted by compatriots, or by other girls of her own age or slightly older, otherwise choosing stories on the basis of being interested by their titles. However, some four months after first visiting KidPub, S readily agreed to a story she had written in an exercise book for her teacher at school being transcribed by her mentor and submitted by email for publication. It was a story of which she was proud, and it had been to some degree influenced by Russell Hoban's classic "The Mouse and His Child" which she had read in the previous fortnight. She dictated an "about the author" endpiece specifically for publication on the site, but otherwise did not wish to revise or redraft the story for WWW publication. In composing this endpiece, she was strongly influenced by similar endpieces she had previously written for handwritten and desktop published versions of stories written in earlier years.

S's writing mentor was doubtful that WWW publication would have any positive effect in making her less of a reluctant writer, although this was the purpose of seeking to involve her in publishing on the Web. Having noticed the degree of interest S had shown in the number of readers accessing individual stories, her writing mentor hypothesized that she would be positively motivated by having a large number of readers access her story. Action was taken by the mentor to publicize the URL for her story once online by circulating details to closed email groups of which she was a member. In this way, the mentor took on another of the roles associated with real world publishing, that of publicizing the publication. Unlike a real world editorial publicist, however, S's mentor did not share with S that she had taken the action,

S was asked to predict the number of readers who would access her story on its first day online. Having seen from her own readings of other children's stories on KidPub that a typical number of readers on the first day would be of the order of 4-12, S predicted that there would be ten or eleven. By the end of the first day, 39 accesses had been logged, to S's surprise and evident delight. Over the next days, S was intensely interested in visiting and revisiting her new story Website, in some part to enjoy the sight of her story in its professionally produced Web format, but primarily in order to see how many more readers had read her story (the number had risen to over 100 within one week). Thus in this first stage, S's interest in the publication was centred on predicting and tracking the number of readers logged at the site from day to day. This initial phase of interest was thus crucially tied to the availability of an onscreen logging facility onsite.

An unpredictedly large number of reader responses to the story began to arrive by email, posing a problem for the child author and her writing mentor. This is a previously unconsidered aspect of the outcomes of encouraging children to write for real audiences. In real world commercial publishing in print, only the most outstandingly successful authors receive substantial responses from readers in the form of direct mail. A world class author selling in the hundreds of thousands of print copies may receive some dozens of letters per week, The analogous hand made and desktop published books produced by children are highly unlikely to receive written responses, since the audiences for the most part are in direct contact with the writers. But with WWW, any given publication is instantly available to many millions of individuals worldwide. Many-to-one communication from otherwise unknown readers to a writer is greatly facilitated by the direct email link to the individual author which sites such as KidPub carry. It is the easiest of tasks to send an instant written response to the author.

Ten email responses to S's story arrived within the first two days of its publication, and within a month over twenty had been received, some by surface mail. Responses came mainly from the US, but also from Canada, France, England and Australia. Most of the messages congratulated the writer on her story and also made some sort of request which presumed that she would reply.

The publication of the story on WWW was visibly having the desired effect of making S feel very positive about her writing and intrigued and interested by others' responses to it. She was almost as interested in the range of places her audience was writing from as she was in the number of daily accesses to her story Webpage. However, her mentor's intention of helping S become a less reluctant writer was compromised by the prospect of successful publication bringing with it the potentially daunting task of drafting large numbers of responses.

Another previously undiscussed facet of the commercial publishing metaphor provided an unexpectedly fruitful solution. The mentor explained to S that very successful authors have secretaries who help them answer the letters they receive from readers. With S's agreement, the mentor took on this role, helping S look through each printed out email in turn, and encouraging S to identify specific requests for information or response. S then dictated a response which was transcribed and emailed by the mentor.

S had had little experience of writing letters for either "real" audiences or in classroom exercises, beyond a small number of thank you letters and holiday postcards. She initially dictated very brief answers to her email correspondence, generating a three stage of message-answer-final message, of which the following was typical

Hi,

I enjoyed reading your story. I will tell it to my 4 year old son

tonight at bedtime. I have a daughter, who is 11, a little older than

you and we live in Philadelphia. Keep writing stories. It is fun to

pass them on. Lucas calls it "stories from your mind"....the best kind.

Thanks for sharing your story with us.

Dear -

Thank you for writing to me. I am glad to hear you like the story. I hope your son likes it. Could you email me back to say if he did?

S

Hi,

Lucas liked your story but said it was "too short". This is probably because

he likes "long" bedtime stories so he can stay up later.

At the end he said that the gecko and mouse had become friends. He especially

liked that part.

Keep up your writing!

However, as S dictated each response, she visibly developed increasing control of the letter form, and began to engage in a much greater level of interactive dialogue. This is the last email exchange of the first sequence of eight emails which she dictated:

S,

I've just read your story on en-garde.com/kidpub

and it was a lot of fun. I was also really scared

when I got to the part about the rattlesnake! Really!

I hate snakes!

I'm glad the story ended the way it did! Can you tell me

what a fruiting tree is? Is it a tree with fruits? Or a

special tree? I love plants and trees and I don't know this

one. What does it look like?

My cat, Whisky, says hello too! She often sits on my lap

when I'm at the computer. Do your cats do that too?

Writing to you from France,

All the best,

L

Dear L

Thank you for writing me such a nice email.

I'm glad you found the story fun. You say you hate snakes. Why do you

hate them? I actually quite like snakes.

Yes, a fruiting tree is a tree with fruits. Try to imagine a fig tree with figs.

I like the sound of your cat. Only one of our cats is very interested in computers. He is the male cat. His name is Pashta. Last night, when we were looking at emails, Pashta jumped on our shoulders, slid onto my mum's lap and started messing up the computer by pressing these different buttons. Then he settled down and stared at the screen.

Where in France do you live? We have been to the Charente. We have also been to Brittany. You may not be familiar with them, but they are both very beautiful.

best wishes to you and your cat.

S

Dear S,

And thank you for such a nice email, too!

You asked why I hate snakes -- simply I'm afraid of them! Where I grew up there were Rattlesnakes and in Florida other poisonous snakes. I'd just as soon stay away from them!

The fruiting tree is such a nice expression. We have lots of trees in our garden, but only apple and pears. I'm not sure I've ever seen a fig tree -- maybe in Tunisia or Morocco... don't remember!

Ah, my dear cat. Whisky is a black cat but a good luck cat. She's my cyberkitty! Sounds like Pashta and Whisky would be good friends. Whisky usually jumps over the keyboard -- I haven't been able to teach her to type yet.

I live in a village called Boullay les Troux which is 35 kms southwest of Paris. I've been to Brittany, but don't know the Charente region. France is a beautiful country almost every place you visit.

Best of luck with your writing and publishing on the web.

Thanks for taking the time to write and hello to your cats from mine!

L

After the first sequence of eight emails, S's time was taken up with other demands. It was almost a month before she and her writing mentor were able to deal with the next batch, making some nineteen responses in total. On both the first batch and the second batch, she and the mentor read through each email and then S dictated her response before proceeding to the next email. On the second round, S initially reverted to the more brief and formal answers of her first few email responses, again showing development of more elaborated interactive replies as she dictated successive answers.

S now began to think of pieces she was writing at school as potentially being suitable for a worldwide audience via WWW. However, she moved house, involving a change of school, and it was two months before she had available a further piece which she wanted published via KidPub. This was a poem on a subject, friendship, chosen by her teacher and composed in writing at school. The poem was transcribed and emailed to the KidPub site by her mentor, who again circulated details via email lists.

S again showed the same pattern of interest as before, initially focussing on the number of accesses to the Webpage with the poem on it, but also now asking if there were emails for her. The number of emails was not as high as for her previous story, but it was again worldwide, with responses coming from Australia and Singapore as well as the countries previously referred to. The same pattern was again observed as S dictated to her mentor responses to her readers. Her first dictated response was to L, her correspondent of the first exchange, who sent her a lengthy email praising various aspects of the poem and asking her if she would like her poem to be chosen as a site of the month for the Web page which L maintained, which is a list of resources for teachers of English as a Foreign Language. S's response was much more limited than her first email to L:

Dear L

Thank you very much for writing to me again.

I am glad you like my poem so much. Thank you for telling me about the typing error. It has now been corrected.

I would be happy for you to put me on your Editor's choice for March 1996.

Sorry this took so long to get to you.

best wishes for you and your computer.

S

However, S's next response, dictated within half an hour of the previous response, showed her as having rapidly retrieved the level of skill and interactivity shown by the last response of her first sequence, and gone well beyond it:

Dear R

Thank you very much for writing to me about my poem.

KidPub is a Website where children of any age can put any story or poem on an actual page of their own. I am not sure if someone of your age could put a story or poem on KidPub, but perhaps your grandaughters and great-neice would like to put some of their stories on KidPub. Please do not say that you are hopeless with computers. You are probably just not used to using them, and you will learn which buttons to press, when and where.

I'm also very glad that you like my poem and story. I might be going to write more. I am thinking of setting up my own home page, with all my stories and pictures on it.

best wishes to you, your two grandaughters, your great niece and your computer.

love from

S

By this stage, then, S sees herself as an author who makes choices, and who is potentially someone who can set up her own home page, which will include all her writings and artwork. This represents a decisive change in her self image, and her mastery of the letter form also shows that she has made major gains in terms of her ability to handle the epistolary dialogue, including her own decisions on the selection of which elements in her correspondents' texts she responds to, and in what order. She clearly seeks to take control of the dialogue, reassuring this senior adult about her self doubts about her computer skills.

The very positive developments in S's view of herself as a writer and in her skills and abilities to plan future writing accomplishments have been accomplished very substantially through the agency of a mentor who was able to devise strategies which ensured that the process of getting stories published and responding to readers was experienced positively by S.

The KidPub site design, with its use of page counters, individual Web pages and individual email response links has also been crucial in enabling S to publish and engage with a worldwide audience. Both S's mentor and KidPub have adopted strategies which leave decisions about design, presentation and the mechanics of Web site creation in the hands of adults; this is essentially an exercise in computer mediated communication rather than in the use of Web design and skills (such as html coding) by children themselves. KidPub is unusual in having few images on its Website, and in adopting a relatively conservative approach compared with many sites directed at children, which frequently image-led, and adopt a style of high-colour, and quirky graphic devices typical of comic book design. Such sites frequently are slow to load because of the degree of graphic complexity they contain. S's response suggests that the professional, adult-looking appearances of KidPub's Web pages is partly what gives them their appeal to the children who publish on them. However, as she begins to envisage herself as owning and generating her own home page, her model is not so much KidPub as other children's Web pages. Most of the latter are equally the outcome of adult rather than child design and control, which is hardly surprising given the present complexity of html production. However, new software being developed by the major commercial software producers is likely to make Web page design by children muchmore of a reality. It remains to be seen whether such producers can resist the easy and obvious solutions of modelling design for children's Web page on the cut-and-paste comic book model, and offer a wider range of approaches which are easy to use for children but which are closer to those of real world publication for adults.

Teachers are still barely able to make available to children the best of desktop publishing and wordprocessing facilities as they now exist. The advent of WWW offers children previously undreamt of access to real audiences worldwide. Teaching of design and communication facilities for Web use is not as yet on the agenda of writing courses for children. It is time for that agenda to be reconsidered.

Web Conceptual Space

Robert E. Kent
School of EECS
Washington State University
USA
rekent@eecs.wsu.edu

Christian Neuss
Fachgebiet Praktische Informatik
Technische Hochschule Darmstadt
Germany
neuss@isa.informatik.th-darmstadt.de

Abstract: In this paper various tools and techniques from the mathematical theory of conceptual knowledge processing [Wille 1982] [Wille 1992] are applied to the conceptual universe of the World Wide Web. For the efficient and economical management of conceptual knowledge, the new idea of conceptual space is introduced. For the conceptual empowerment of the Web user, a new technique called conceptual browsing [Kent & Neuss 1996a] is advocated. Conceptual browsing, which browses over conceptual space along conceptual links, is dual mode (extentional versus intentional) and dual scope (global versus local).

Introduction

Discovery of relevant resources is still one of the biggest problem in the continuously growing infrastructure of the World Wide Web. In order to allow the Web to become a true electronic library, it needs a service that identifies and locates documents belonging to a given category, based on descriptive document attributes such as author, title, or subject. In traditional library science, resources are identified through catalog information, which is organized as a hierarchy of restricted subject categories. Searches are conducted upon this explicitly encoded descriptive information, and can easily be supported with relatively simple retrieval programs. Although traditional library science maintains indexes manually, the dynamic nature of electronic documents requires a certain automation of this task. Not only does the pure vastness of the document space forbid manual indexing, but many of the electronic documents are also subject to frequent change. Currently, the popular searchable indexes of World Wide Web resource information are based upon Web robots: programs that connect to a remote site and recursively retrieve documents over the Network, usually in order to build an index. Since the HTML markup language provides relatively little support for encoding document structure and semantic information [Nicol 1995], automatic tools have to search the document text and employ heuristic methods in order to judge document relevance.

In order to address this problem, recent research aims at providing self-identifying document structure [Barnard et al. 1995], standards for the representation of metadata [Weibel et al. 1995] [Bowman & Schwartz 1996], and scaleable, distributed directory services for locating information resources such as Harvest [Bowman et al. 1994a] or the Synopsis File System [Bowman 1996]. Provided that sufficient descriptive information about documents is available for resource discovery tools, the question arises how the information acquired can best be digested and manipulated. One approach is to employ a graphical metaphor to represent information structures, and provide users with a map-like view of the underlying information space [Robertson 1993]. These information visualization techniques can successfully be applied to the World Wide Web [Mukherja & Foley 1995]. However, creation of an effective view is difficult, as the available descriptive information lacks many useful semantic attributes [ibid.]. In general, visualization involves a two- or three-dimensional projection from a multidimensional attribute space, and thus requires an interpretation of the raw attribute data.

Conceptual analysis of Web metadata [Kent & Neuss 1994] [Neuss & Kent 1995] provides a formal mechanism for the automatic categorization and classification of documents. By applying *conceptual scales* to metadata, faceted conceptual spaces can be constructed. In these concept spaces, various properties collocate Web document-like objects according to common characteristics. Conceptual scaling is an interpretative act with a highly interactive user interface component. This paper advocates a method called *conceptual browsing* as a navigation and structural browsing technique. It is based on the notion of "conceptual distance," a formal similarity metric for documents. The algorithmic definition of similarity allows for an incremental browsing of the conceptual vicinity of a document or a class of documents.

Conceptual Space

Conceptual Knowledge Processing

Conceptual knowledge processing [Wille 1982] [Ganter & Wille 1989] is a relatively new discipline arising out of the mathematical theory of lattices and the calculus of binary relations. It is closely related to the areas of knowledge representation in computer science and cognitive psychology. Conceptual knowledge processing provides for the automatic classification of both knowledge and documents via representation of a user's faculty for interpretation as encoded in conceptual scales. Such conceptual scales correspond to the facets of synthetic classification schemes, such as Ranganathan's Colon classification scheme, in library science. Conceptual knowledge processing uses *objects*, *attributes* and *formal concepts* as its basic constituents. Objects and attributes are connected through has-a incidence relationships, while formal concepts are connected through is-a subtype relationships. In many contexts appropriate for Web resources, the objects are document-like objects and the attributes are properties of those document-like objects which are of interest to the Web user. Conceptual knowledge processing uses formal concepts as its central notion and uses concept lattices as an approach to knowledge representation [Wille 1992]. A formal concept (also called a conceptual class or category) consists of a collection of entities or objects exhibiting one or more common characteristics, traits or attributes. Formal concepts are logically characterized by their extent and intent. The extent of a concept is the aggregate of entities or objects which it includes or denotes. The intent of a concept is the sum of its unique characteristics, traits or attributes, which, taken together, imply the formal concept. The process of subordination of concepts and collocation of objects exhibits a natural order, proceeding top-down from the more generalized concepts with larger extent and smaller intent to the more specialized concepts with smaller extent and larger intent. This is-a relationship is a partial order called generalization-specialization. Concepts with this generalization-specialization ordering form a class hierarchy called a *concept lattice*. This is the formal representation for conceptual knowledge. The use of formal concepts as a conceptual structuring mechanism corresponds to the use of similarity clusters in information retrieval, although foundationally, formal concepts are based more on logical implication rather than a nearness notion.

A *concept space* is a named substructure of a concept lattice. In a precise mathematical sense, a concept space is a "distributed formal concept." The main purpose for the concept space notion is to provide for the organization and customization by the user of their own information space. The representational mechanism of a concept space serves as a firm foundation for the basic paradigms of Internet/intranet resource discovery and wide area information management systems; both organization-navigation and search-retrieval [Kent & Bowman 1995]. The use of conceptual space is a natural outgrowth of the original conceptual knowledge processing approach for structuring and organizing the networked information resources in the World Wide Web [Kent & Neuss 1994]. Conceptual space forms its naming mechanism via conceptual views. A *conceptual view* is a name for a formal concept within a concept lattice. Conceptual views are created by any of the following methods: (1) view definitions in terms of meets and joins of other views, meets of attributes, and joins of objects; and (2) view definitions (*virtual views*) in terms of the meet-vectors of apposition concept spaces or the join-vectors of subposition concept spaces. Virtual views have applications in: (i) the interpretation by conceptual scaling; (ii) the parallel implementation of conceptual knowledge processing, and (iii) collaboration in distributed information systems. Virtual meet views are useful in conceptual scaling when we use various facets of information in order to organize a conceptual space over an information system. For collaboration, we can merge remote public data interpretation, as represented by conceptual scaling, with local private data interpretation. This provides one approach for the private customization and sharing of public data. According to conceptual knowledge processing, knowledge representation is extended by the two techniques of knowledge inference and knowledge acquisition. In conceptual space, knowledge inference is formally represented by: (1) view order closure; (2) view extent/intent relational closure; and (3) inclusion constraint between extent/intent composition and incidence. Knowledge acquisition is formally represented by: (1) any initialization construction from a tree (forest or directed acyclic graph) hierarchy; (2) interactive view definitions; and (3) the structural operations of (a) producting (apposition and subposition) and (b) summing. Conceptual browsing is navigation through conceptual space. Considering the dualistic extent/ intent structure of a concept, it has two temporally disjoint modes; in extentional mode we browse over concepts with respect to their extents, whereas in intentional mode we use conceptual intents in a strictly dual manner. Conceptual browsing has both a local and a global scope. Browsing in the global scope means browsing over the entire concept lattice, whereas browsing in a local scope means browsing over a local neighborhood concept lattice. There are two dual senses or modes for the idea of a local neighborhood of a formal concept within a concept lattice. The extentional neighborhood of a concept regards the concept as an attribute; it fuses the intent of the concept as a collective attribute and distributes the extent downward over a local neighborhood lattice. The intentional neighborhood is strictly dual. Conceptual browsing is based upon a notion of conceptual similarity. The extentional similarity, which is used in the extentional mode, is a measure of the similarity of any two concepts according to their common extent cardinality. Again, the intentional similarity is strictly dual.

WWW Conceptual Space

In order to illustrate the ideas for organizing conceptual knowledge which have been introduced in this paper, we here discuss a familiar conceptual knowledge universe, the universe of World Wide Web conferences; in particular, the concept space of the proceedings documentation of WWW'95 (*http://www.igd.fhg.de/www95.html)*. At the top level, these proceedings consist of papers, posters, tutorial notes/slides, and workshop proceedings. The WWW'95 proceedings encompass a diverse range of topics (e.g., WWW protocol enhancements, resource discovery, retrieval, web security, etc.). But in order to provide an effective user-interface, the metadata from the Web documentation needs to be conceptually scaled according to the needs and interests of the user. Such conceptual scaling can be accomplished by user-defined conceptual views.

Obtaining meaningful descriptive document data is a crucial step. Descriptive information can be encoded in document markup. The online version of the proceedings of WWW'95 uses the <META> tag in HTML, in order to augment documents with metadata such as title, names of authors, or keywords. [Tab. 1] shows a document header with embedded metadata. The information stored as attribute/value pairs can easily be extracted, and postprocessed to provide a machine parsable interchange format. As interchange format [Tab. 2], currently we are using a BibTeX-derived format compatible with the Harvest Summary Object Interchange Format (SOIF) [Bowman et al. 1994a] and the Synopsis File System synopsis [Bowman 1996]. Later we intend to structure metadata in terms of the entity-relationship data model.

```
<HTML>
<HEAD>
<TITLE>WWW'95: Towards an Intelligent Publishing
Environment</TITLE>
<META NAME="Topics" CONTENT="Living Documents">
<META NAME="Keyword" CONTENT="Internet tools,
electronic publishing, hyperlink databases">
<META NAME="Author" CONTENT="James E. Pitkow, R. Kipp
Jones">
</HEAD>
</HTML>
```

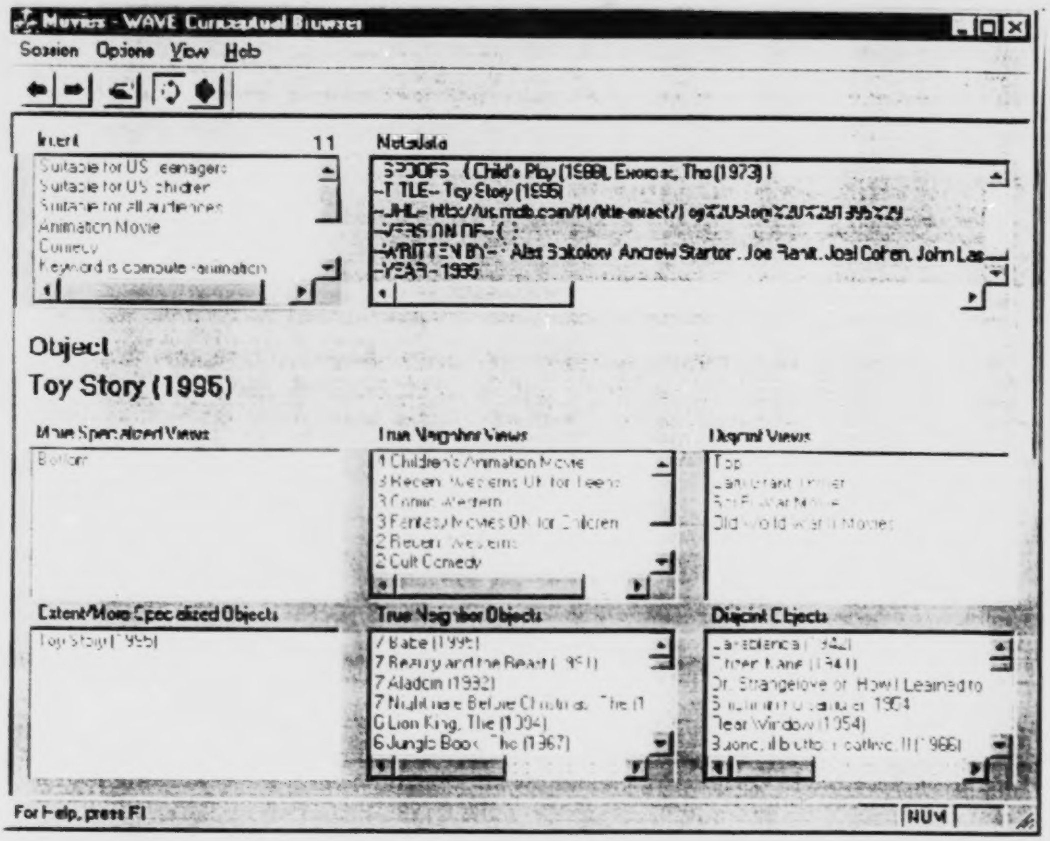
Table 1: Metadata in HTML Documents

```
@Paper {
UID = pap72,
URI =
"/www/www95/proceedings/papers/72/publish/publishing.html",
TITLE = "WWW'95: Towards an Intelligent Publishing
Environment",
SIZE = "33447",
SESSION = "Living Documents",
KEYWORDS = {"Internet tools", "electronic publishing",
"hyperlink databases"}
AUTHORS = {"James E. Pitkow", "R. Kipp Jones"}
}
```

Table 2: BibTeX-based Interchange Format for Metadata

Navigating Conceptual Space: An Application

The ideas discussed in this paper have been implemented in the Networked Information Discovery and Retrieval (NIDR) system WAVE (http://wave.eecs.wsu.edu/), which is being developed in a project entitled "Creating a WAVE." This project aims at providing an answer to the question "what is the appropriate architecture for a digital library?". While the current prototype of the WAVE conceptual browser (a downloadable copy is available from the WAVE Web site) is a standalone application running under Windows, it will later be made accessible as a Java-based application.



Suitable for US teenagers
Suitable for US children
Suitable for all audiences
Animation Movie
Comedy
Keyword is computer-animation

SPOOFS {Child's Play (1988), Exorcist, The (1973)}
TITLE-- Toy Story (1995)
WRITTEN BY-- Alec Sokolow Andrew Stanton, Joe Ranft, Joel Cohen, John Las
YEAR-- 1995

Toy Story (1995)

Bottom

1 Children's Animation Movie
3 Recent Westerns OK for Teens
3 Comic Western
3 Fantasy Movies OK for Children
2 Recent Westerns
2 Cult Comedy

Top
Old World War II Movies

Toy Story (1995)

7 Babe (1995)
7 Beauty and the Beast (1991)
7 Aladdin (1992)
7 Nightmare Before Christmas, The (1
6 Lion King, The (1994)
6 Jungle Book, The (1967)

Casablanca (1942)
Citizen Kane (1941)
Dr. Strangelove or: How I Learned to
Rear Window (1954)

Table 3: **The Movie:'Toy Story' page (screen shot)**

[Tab. 3] demonstrates conceptual browsing with the WAVE system over movie data. By accessing the Internet Movie Database (http://us.imdb.com/), data relating to movies has been automatically extracted and pre-processed to produce a base format suitable for use in the WAVE conceptual browser. To illustrate the process of conceptual browsing over a movie concept space, we chose a small database of 250 movies. You begin with an empty browser. You can choose the application which you want to browse over (in this case movies) from a menu of applications. You next select facets of interest to you from the movie application. Movie facets might include: Actor, Actress, Certification, Company Produced By, Genre, Keyword, Length, Period and Producer. Suppose you chose the 5 facets: Actor, Actress, Certification, Length and Period. The initial conceptual similarity browse defaults to the extentional mode and the 'Top' view for all applications. At this point, suppose we want to explore the recent movie 'Toy Story'. We can transition from the Top conceptual view to any movie object, such as 'Toy Story', by selecting it in the extent list-box and pressing *Return*, or simply double-clicking on the movie. After transitioning, the current conceptual state is the 'Toy Story' 'movie object and the current conceptual display is the Toy Story' page [Tab. 4].

There are three main components in this conceptual display: the formal concept display, the conceptual similarity display, and the metadata display.

- The first column represents the *formal concept display*: the top list-box contains the intent of the formal concept generated by the movie object 'Toy Story'; next down is the title of the movie; third is the list-box of more specialized views; and finally at the bottom is the list-box containing the extent or more specialized objects.
- The bottom two rows represent the *conceptual similarity display*. Since the current conceptual

BEST COPY AVAILABLE

state is a movie object, the conceptual similarity is computed in the intentional mode. For example, the similarity between 'Toy Story' and the 'The Lion King' is 6, meaning that 'Toy Story' and 'The Lion King' share 6 attributes in common. From the 'Toy Story' page you can see that of the 250 movies which are in the database, 'Babe', 'Beauty and the Beast', 'Aladdin' and 'Nightmare before Christmas' are the most similar to 'Toy Story' (they share 7 out of 11 attributes). On the other hand, movies which have no attributes in common with 'Toy Story' include 'Casablanca' and 'Citizen Kane', etc.

- The text-box in the top right represents the *metadata display*. In order to get a generic format for any application and any object-type, we intend to interpret metadata as the description of a *bibliographic entity* or *work*. A generic form for this is provided by the ISBD (International Standard Bibliographic Description) consisting of 8 areas: (1) title and statement of responsibility; (2) edition; (3) material (or type of publication) specific details; (4) publication, distribution etc.; (5) physical description; (6) series; (7) note; and (8) standard number and terms of availability. The standard number corresponds to the URL field in [Tab.4], which points to the actual digital object.

Summary and Future Work

The WAVE navigator implementation has currently an off-line demonstration conceptual browser running over applications for movies, music, etc. We are currently experimenting with functionality and usability. During the next year we will be experimenting with the scaleability issue, both in the sense of larger databases and in the sense of an on-line client/server architecture. Both on-going and future work can be discussed in terms of three processes for the conceptualization of networked information resources: metadata abstraction, conceptual scaling, and conceptual linkage [Kent & Neuss 1996a]. The first process, *metadata abstraction* (also called summarization, gathering, or synopsizing), is often implemented as the first (back) component of a NIDR system. An important example of a metadata abstraction processor is the gatherer component of the Harvest system [Bowman et al. 1994a]. The third process, *conceptual linkage*, is now being implemented as the first phase of the WAVE system development. This phase replaces the broker indexing component of the Harvest system, extending broker capabilities by adding dynamic and customizable knowledge organization techniques. It is being used for interactive information analysis and browsing guidance during exploratory search by client Web browsers over a community's information space. The second process, *conceptual scaling*, will be implemented as the second phase of WAVE system development. This phase will represent the process of faceted analysis which occurs in library science classification. It also corresponds to the design of user interest profiles in current awareness services [Rowley 1987].

References

[Barnard et al. 1995] Barnard, D., Burnard, L., DeRose, S., Durand, D., & Sperberg-McQueen, C. (1995). Lessons for the World Wide Web from the Text Encoding Initiative. Proceedings of the Fourth International World Wide Web Conference, December 1995, Boston. 349-357.

[Bowman 1996] Bowman, C.M., (1996). The Synopsis File System: Information Management in Wide-Area File Systems. Presented at the Decorum'96 conference, February 1996, Tarpon Springs, Florida.

[Bowman et al. 1994a] Bowman, C.M., Danzig, P., Hardy, D., Manber, U., & Schwartz, M (1994). Harvest: A scaleable, customizable discovery and access system. Technical report CU-CS-732-94, Department of Computer Science, University of Colorado, Boulder. July 1994.

[Bowman & Schwartz 1996] Bowman, C.M., & Schwartz, M. (1996). W3C distributed indexing/searching workshop report. Technical report, http://www.w3.org/pub/WWW/Search/9605-Indexing-Workshop/, the World Wide Web Consortium (W3C), Cambridge, Massachusetts. May 28-19, 1996.

[Ganter & Wille 1989] Ganter, B. & Wille, R. (1989). Conceptual scaling. Applications of Combinatorics and Graph Theory in the Biological and Social Sciences, Roberts, F. editor, Springer, New York, 139-167.

[Kent & Bowman 1995] Kent, R.E., & Bowman, M. (1995). Digital libraries, conceptual knowledge systems, and the Nebula interface. White paper, http://www.transarc.com/afs/transarc.com/public/trg/papers/techrep0495.ps, Transarc Corporation. June 1995.

[Kent & Neuss 1994] Kent, R.E., & Neuss, C. (1994). Creating a 3D Web Analysis and Visualization Environment. Computer Networks and ISDN Systems, 28, 109-117. Also appears in the Electronic proceedings of the second international World Wide Web (WWW) conference'94, Mosaic and the Web.

[Kent & Neuss 1996a] Kent, R.E., & Neuss, C. (1996). Conceptual analysis of hypertext. Advances in hypertext. Lecture notes in computer science series, Nicholas, C. & Mayfield, J. editors, Springer-Verlag.

[Mukherja & Foley 1995] Mukherja, S. & Foley, J.D. (1995). Visualizing the World-Wide Web with the navigational view builder. Computer Networks and ISDN Systems, 27, 1075-1087. Also appears in the Electronic proceedings of the third international World Wide Web (WWW) conference'95.

[Neuss & Kent 1995] Neuss, C., & Kent, R.E. (1995). Conceptual analysis of resource meta-information. Computer Networks and ISDN Systems, 27, 973-984. Also appears in the Electronic proceedings of the third international World Wide Web (WWW) conference'95.

[Nicol 1995] Nicol, G.T. (1995). Dynaweb - integrating large SGM repositories and the WWW. Proceedings of the fourth international World Wide Web conference, Boston, 549-558. December 1995.

[Robertson 1993] Robertson, G.G. (1993). Information visualizing using 3D interactive animation. Communications of the ACM, 36 (4), April 1993.

[Rowley 1987] Rowley, J.E. (1987). Organising knowledge: An introduction to information retrieval. Aldershot, Hants, England: Gower.

[Weibel et al. 1995] Weibel, S., Godby, J., & Miller, E. (1995). OCLC/NCSA metadata workshop report. Technical report, http://www.oclc.org:5046/conferences/metadata/dublin_core_report.html, OCLC, Dublin, Ohio. March 1995.

[Wille 1982] Wille, R. (1982). Restructuring lattice theory: An approach based on hierarchies of concepts. Ordered Sets, Rival, I. editor, 445-470, Dordrecht-Boston: Reidel.

[Wille 1992] Wille, R. (1992). Concept lattices and conceptual knowledge systems. Computers and mathematics with applications, 23, 493-522.

Acknowledgments

This research was supported by a grant from Intel Corporation. Many thanks to Abijah Perkins and Jeff Olson, who are team members on the project "Creating a WAVE."

An Information Broker for Adaptive Distributed Resource Discovery Service

Arkadi Kosmynin
Division of Information Technology
Commonwealth Scientific and Industrial Research Organisation
Australia
Arkadi.Kosmynin@mel.dit.csiro.au

Abstract: In the paper we describe What'sHot: a resource discovery system under development that uses active agents (information brokers) to organise distributed dynamic adaptive indexes of resources. What'sHot utilises the information, that is available as a side effect of human search and retrieval activity, to maintain indexes of *valuable* resources. The system is designed to provide scalable index service on the Internet and to facilitate co-operative information filtering. This paper discusses design and philosophy of What'sHot.

Introduction

Rapid growth of the Internet and WWW [December 94] produces an urgent need for tools which assist people in search for resources on the Net. In this paper we present a simple idea: a decentralised, distributed and changing environment, such as the Internet, needs a corresponding dynamic, distributed, adaptive index service.

The existing resource discovery systems can be divided into two major categories in relation to information suppliers and users. The first group consists of systems which build and optimise indexes and directories of available resources. Large centralised indexes (Alta Vista, InfoSeek, Excite, ALIWEB [Koster 94], WebCrawler [Pinkerton 94]), directories (Yahoo, WebDirectory, McKinley Directory), and distributed indexing systems (WHOIS++ [Weider 96], Harvest [Bowman 95], Discover [Sheldon 95]) belong to this group.

The second group consists of systems which try to actively assist users in their search process. These systems organise search process using information about users' preferences and interests. Intelligent agents (Jasper [Davies 95], InfoScout [Voorhees 94], WEBDOGGIE) and user interfaces/smart query routers (BabyOIL [Ianella 95]) form this group.

The disadvantage of the current situation is that these two groups are isolated in the index building/optimising stage. Given that adequate resource discovery indexes must be dynamic to adapt to the Internet's changes, the information obtained from users should be utilised to control those dynamics and to optimise the indexes.

The Internet is a product of collaborative efforts of millions of people and there is a huge amount of human activity happening every single moment on the Net. This activity has a potential to be transparently used to make the Net's anarchy self-organising. What'sHot uses active agents to organise distributed dynamic adaptive indexes of meta-information. Units of meta-information (abstracts) travel from index to index, automatically finding their way to a consumer. It is scalable, economic, and adapts as the environment changes (change of resources available on the Net and changes in people's interests). It uses the joint expertise of the Net community to assist people in search for *valuable* resources. It has a number of other distinguishing features which we describe in this paper.

What'sHot

Our approach is based on automatic distribution and replication of meta-information (abstracts of resources).

By the "abstract" we mean a unit of meta-information, a small (within 1-2K) description of a resource. An abstract has to contain at least two components: a short description of resource to judge on the relevance to a query being processed and some pointer to the resource (such as URL or URN) to locate the resource at retrieval stage. WHOIS++ templates and Harvest SOIF templates are examples of an abstract. Other information, such as abstract's expire date, its source, popularity rating, may be stored with the abstract.

Using the abstracts instead of full text documents makes it technologically feasible to pass, replicate, index and search the resource discovery meta-information because it decreases index size and network traffic. It also provides a uniform approach other types of resources, such as pictures and software packages.

To achieve aims of distributed indexing and query processing, we could deploy a network of index brokers and allow abstracts to travel from index to index and stay longer where and while they are popular. The more popular the abstract is, the faster and further it would travel and the more replicas would exist.

Observations and Assumptions

A number of observations led to this approach and which suggest that it can achieve its aims:

1. When one is looking for information, it is very likely that others have already looked for this information. This is especially true in a specialised environment, where people are working in the same area and require similar information. This observation is supported by the success of resource caching [O'Callaghan 95] and mirroring. Also analysis of log files of search engines has shown that there are small number "hot" topics which constitute a large part of the total number of queries.

2. In many cases, users would be satisfied with any information on the search subject because they can use this information to find more links or references.

3. A large amount of published material on the Internet has a very low value, and there is a strong need for expert selection, not for search returning anything matching query terms.

Because of these reasons, the News, for example, is successfully used as a resource discovery system, although it was not designed for that purpose. In our experience, it is a good idea to try the news FAQs as the first point when starting looking for material in a new area.

In What'sHot, we assume that we can obtain abstracts of resources from publishers, or derive abstracts automatically from the resources, or get them from existing systems, like Harvest. We also assume that having retrieved and reviewed a document, people are willing to give a "yes/no" feedback (vote) on the document value and relevance to the query. This is not mandatory, however. We can judge on the value of resource based on the number of users who retrieved it, but an explicit vote is a better criteria.

What What'sHot Looks Like For Users

There is no essential change in the current search process. Instead of contacting InfoSeek or Lycos, people contact a (local, in most cases) What'sHot broker and receive a search form. After submitting a query, they receive a result form similar to the one you would receive from a conventional search engine. What'sHot result form, however, has a check box for each abstract. There is also a "Submit" button on the form. The users check the check box only if they liked the resource. They press the "Submit" button to send the feedback to the broker.

A user may send their user profile (description of interests) to a broker and (possibly, for a small fee) from time to time receive e-mail messages about newly available *good* resources that may be of interest.

We realise that there are users (possibly, many), who do not bother checking boxes and replying. The system just ignores them. This means that the system does not cache information that is interesting to these people. In the worst case, we can make a rough judgement on the value of resource transparently for users based on how often it has been retrieved by users.

We feel it would be a good idea to also include at least "Could not get" check box and "Submit" button into result forms of other search engines. This way would enable the search engines to clean their indexes.

How It Works

Each broker consists of a database and software that performs following functions:

1. Maintain, index and search the local cache of the abstracts..

2. Process incoming queries locally and send results back to clients.

3. If a query can not be processed locally or the user is not satisfied with the results of local processing, make decisions on routing the query to other sources of meta-information.

4. Make decisions on caching (replication) of remote query results (which are sets of abstracts) locally.

5. Publish (make available for queries) the abstracts of local and/or authorised users.

6. Periodically recommend the most popular abstracts to other brokers that might be interested in getting a copy of the abstract for replication.

7. Make decisions on caching locally abstracts recommended by other brokers.

Data

The way we build the broker's cache is similar to what happens in the process of human learning. We (human beens) collect relatively small bits of knowledge about every possible topic we face in our life, from road traffic rules to rumours about crimes. The more it is interesting for us, the more often we use this knowledge, the longer we remember it. We also have knowledge (often vague) about possible sources where we can get more information on a given topic, if necessary. This part of memory is our knowledge cache and knowledge map.

Most of people are experts in a narrow area. If they are recognised as the experts, their expertise is used by other people directly or indirectly. If one is a known expert in some area, we ask them for information on this area and sometime we suggest a new information for their judgement or if we want this information to be known to other people through them or just to contribute to their expertise.

The broker's database contains three groups of abstracts.

- Abstracts which are popular among the local users and form a corporate user profile. We call this group the "wide view group". It is knowledge cache and knowledge map of the broker.

- A specialised group of abstracts on some topic, prevailing in the wide view group. We call this group the "expertise" group.

- Abstracts that are published by local users. We call this group the "local group".

Additional information is stored for each abstract. This includes expiry date, date last used (for resource retrievai or query routing), source broker, popularity rating, cost of retrieval, user profiles.

It makes sense to store abstracts of the wide view group because the information that is interesting for one user within the organisation, is very likely to be interesting for the others. Second, this group of abstracts forms a knowledge map of the outside world of information sources. It contains a dynamic set of abstracts covering a wide range of topics. This information is used to choose the best sources for given search topic.

Storage of the expertise group makes sense because we need to concentrate knowledge of a topic as much as possible to reduce need for query distribution. In other words, we have to achieve specialisation of a broker in some topic. If we do this, only a few brokers need to be searched for a particular topic, or a combination of topics, instead of the millions of brokers as in the case of dispersed information.

Finally, users have the right to make their local broker store and publish their local abstracts, disregarding the popularity of the resources. However, the broker should not supposed to recommend an abstract for replication, unless the abstract is popular.

Work of a Broker in Normal Regime

The normal regime describes work of a broker in the case when there is an existing network of interacting brokers and the broker we are considering has been working long enough to adapt to the environment.

Publishing

Publishers supply abstracts of resources being published to the brokers to which the publishers have publishing access. The brokers store these abstracts. After they have been stored and indexed, the abstracts are available for all the queries coming to the brokers. However, an abstract is not recommended for replication on other brokers until it becomes popular.

Brokers periodically check which abstracts have collected enough votes for replication but have not been recommended for replication. For each of the abstracts from this set, the broker tries to determine the best broker to replicate it on using the FindSimilar() function. This function takes a text A and a set of texts B and returns the texts from B that are close to A. So, the broker performs a chain of transformations:

FindSimilar(abstract A to recommend, wide view + expertise groups) -> closest abstracts -> their source brokers -> the best broker B to recommend the abstract A to. The broker sends a replication request to broker B. If B decides not to replicate the abstract A, B repeats the procedure and refers A to other brokers. More than one broker may be selected to recommend an abstract to.

Search

The user submits a query to their preferred broker. The broker uses the function FindSimilar(query, all abstracts) to get a result set. If the result set is substantially large and ranking scores are high, the set is sent back to the user. If the user is not satisfied with the set received, they have option to refine the query terms or send a "more" request. Having received the "more" request, the brokers that processed this query consider further distribution of the query if they do not have any more relevant abstracts to send. This enables users to expand searches more and more if not satisfied with the query results. If the broker failed to find results within the network of brokers, it can submit the query to an external source, such as InfoSeek.

To route a query, broker performs a chain of transformations:

FindSimilar(query, wide view + expertise groups) -> closest abstracts -> their source brokers -> the best broker to route the query.

Caching

The cache manager is a component of the broker software that is responsible for implementing the cache policy. It makes decisions on storing a copy of abstract in the broker. It takes into account the local popularity of the abstract, relevance to the specialisation area of the broker and whether this abstract is easy to get from another broker if needed. It replicates an abstract if retrieval of the abstract is expensive (in terms of communication resources) or if it corresponds to the area of broker specialisation. The cache manager also decides which abstracts have to be discarded from the cache to free space for a new abstract. The cache policy is regulated by setup parameters, such as portion of the cache to allocate for the wide view group and portion for the expertise group, the maximal value of retrieval cost before replicating an abstract, the minimal number of votes collected by an abstract to be replicated, etc. These values might also be used as weights for components of decision criteria.

These simple parameters significantly affect broker behaviour. For example, by maximising the expertise partition, we can achieve more value of the broker as an expert. By increasing the value of the retrieval cost necessary for replication, we reduce the number of brokers replicating a given abstract, thus, potentially reducing whole amount of the "expert" brokers with duplicated expertise.

All abstracts that are obtained as query results and recognised as valuable after presentation to users, are cached locally, at least in the wide view group, if the cost of their retrieval is higher than some threshold. The LRU (Least Recently Used) algorithm is used to update the cache. If there are many similar abstracts to a new one in the expertise group, and the ranking scores are high enough, the new abstract is included into the expertise group.

Adding a New Broker

A broker consists of a software and a database. The parts that differ one broker from another one are setup parameters and the database. Setup parameters must be provided by the owner of the broker. However, to make a new broker working, we have to supply it with knowledge of the world, which is the content of the database. To do this, we can "transplant" a database from any working broker.

In the worst case, the donor broker is topologically far from the receiver and there is nothing common between the corporate user profile of the owner of the donor and the owner of the receiver. This means that:

1. The database information about the retrieval costs is wrong. This leads to ineffective query routing.

2. The database is irrelevant to the local users. This leads to remote processing of the most of queries.

However, in both cases the situation improves gradually, with the broker updating its cache and, thus, adapting to the changed conditions. To achieve better results in a shorter time, one should borrow the database from as "close" a broker as possible. We mean "close" in topological and corporate user profile aspects.

Another way to start a new broker is to supply it with address of the closest functioning broker for default query routing. In this case, for the first time, most of the queries will be re-routed to that one until the new broker accumulates enough knowledge for independent functioning.

What Can What'sHot Do for Its Users

Suppose only one organisation is running a broker of this type. The broker enables people to transparently share their expert knowledge about the availability and value of the resources, thus reducing the time and cost of searchs for information. The communication cost of a search is reduced because many queries are resolved locally. The broker in this case works as a very large dynamic searchable corporate hot list filled with useful URLs which people can transparently share. People do create such hot lists at the moment, but they do it manually, cannot search them and do not transparently share them.

Suppose that there is a deployed network of the brokers. Given that each broker can maintain information about thousands of other brokers, the overall indexing capacity of this network is enormous. A functioning network composes a kind of serachable space and enables people to publish their materials within this space. This is a remarkable decentralised way of publishing where everyone can publish anything, but only the best materials will be replicated widely and find their way to the mass consumer.

Acknowledgements

The work reported in this paper has been funded in part by the Co-operative Research Centres Program through the Department of the Prime Minister and Cabinet of the Commonwealth Government of Australia.

Conclusions

Our system is currently under development. We believe that this approach can achieve its aims because it is simple and because it harmonically combines and uses a number of ideas that proved to be working: caching, computer viruses, news, reviewing of materials for publication, specialisation in human society.

In conclusion, we would like to highlight distinguishing features of What'sHot:

1. It is easy to use for users.

2. It is transparent for service providers (because it is fully automatic and adaptive) and easy to deploy.

3. It is self-promoting, because both sides (service providers and users) have an interest in it. Service providers are interested in giving access to brokers because the brokers can not function without knowledge that they receive in the process of users' search.

4. It enables people to transparently share their expertise.

5. It is economic because it replicates only valuable information.

6. It is designed to use existing technologies and information sources.

7. It provides a decentralised way for publishing of information, along with existing centralised ones.

8. It provides an opportunity for fair commercial advertising on the Net (because advertising is a type of publishing).

9. It provides opportunity to trade brokers' databases.

10. It is designed to automatically adapt to the Internet changing environment.

11. It is scalable.

Bibliography

[Bowman 95] Bowman, C. M., Danzig, P. B., Hardy, D. R., Manber, U., Schvartz, M. F., Wessels, D. P. (1995). Harvest: A Scalable, Customizable Discovery and Access System. Technical Report CU-CS-732-94. Department of Comuter Science, University of Colorado - Boulder, Colorado.

[Davies 95] Davies, J., Weeks, R., Revett, M. (1995) Jasper: Communicating Information Agents for WWW. The Fourth International World Wide Web Conference, 1995, Boston, Massachusetts.

[December 94] December, J. (1994). Challenges for Web Information Providers. Computer-Mediated

Communication Magazine, Volume 1, Number 6, pp. 8-14.

[Ianella 95] Ianella, R., (1995). Internet Resource Discovery with BabyOIL. AUUG Queensland Summer Technical Conference, 1995, Brisbane, Queensland, Australia.

[Koster 94] Koster, M., (1994). ALIWEB: Archie-Like Indexing in the Web. First International World Wide Web Conference, 1994, Geneva, Switzerland.

[O'Callaghan 95] O'Callaghan, D., A Central Caching Proxy Server for WWW Users at the University of Melbourne. AusWeb'95, 1995, Ballina, NSW, Australia.

[Pinkerton 94] Pinkerton, B., (1994). Finding What People Want: Experiences with the WebCrawler. First International Conference on the World Wide Web, 1994, Geneva, Switzerland.

[Sheldon 95] Sheldon, M. A., Duda, A., Weiss, R., Gifford, D. K. (1995). Discover: A resource Discovery System Based on Content Routing. The Third International World Wide Web Conference, 1995, Darmstadt, Germany.

[Voorhees 94] Voorhees, E. M., (1994). Agent Collaboration as a Resource Discovery Technique. CIKM Workshop on Intelligent Information Agents, 1994, Gaithersburg, Maryland.

[Weider 96] Weider, C. , Fullton, J., Spero, S. (1996). Architecture of the WHOIS++ Index Service. RFC 1913. February, 1996, Available at ftp://ds.internic.net/rfc/rfc1913.txt.

Towards a Novel Networked Learning Environment

Anastasios Koutoumanos (tkout@softlab.ntua.gr) [1]
Nikolaos Papaspyrou (nickie@softlab.ntua.gr) [1]
Simeon Retalis (retal@softlab.ntua.gr) [1]
Hermann Maurer (hmaurer@iicm.tu-graz.ac.at) [2]
Emmanuel Skordalakis (skordala@softlab.ntua.gr) [1]

[1] National Technical University of Athens
Department of Electrical and Computer Engineering
Division of Computer Science
Polytechnioupoli, 15780 Zografou, Greece

[2] Institute of Information Processing and Computer Supported Media
Graz University of Technology
A-8010 Graz, Austria

Abstract: Education can benefit from the use of technology in the teaching process. This has been done in the past using various technologies. In recent years, computers, information systems, multimedia and hypertext systems have been used to develop self-learning environments (CAI, CAL, CBT systems). A drawback of these environments is the lack of interaction to the desired degree. With the use of new technologies like computer networks and networked hypermedia systems, new networked learning environments can be built that alleviate this problem. One such novel Networked Learning Environment is presented in this paper. Its most valuable components are the World Wide Web, the HYPER-G system and the Internet.

1. Introduction

The teaching model that is used in conventional universities is based on the oral presentation of the educational material, in form of lectures, and uses as a supplement exercises and laboratory experiments. Although other techniques have often been suggested (tutorials, cognitive learning, Socratic approach), none of them has yet succeeded in replacing lectures [Young 1995, Ferguson 1975]. The major problems that conventional universities are facing nowadays are:

- Lecture attendance of undergraduate courses, especially courses offered to senior students, decreases. At the National Technical University of Athens, in a course offered in the 9th semester the percentage of attendance is small - usually less than 60%. Some of the main causes are: (1) the fact that many students have already started engaging in professional activities; (2) the prohibitive time that it takes a student to move from his/her home to the university, because of regular heavy traffic; and (3) other occasional causes like health problems, family, personal or social reasons, etc. It should be noted here that in Greece, by tradition, higher education including course books is fee-free and that in most Greek universities lecture attendance is not compulsory.

- It is difficult for students to ask questions and receive answers outside of a lecture's appointed time. It is sometimes difficult to do so during the lectures, because of time restrictions and other personal reasons (anxiety about mistakes, modesty, shyness, etc.). It is also impossible for students to re-attend a lecture.

- The course material changes rapidly and printed course books quickly become obsolete. As a result, the instructor is often forced to provide additional lecture notes and students are often compelled to buy many books using own funds.

- It is generally difficult to locate and access useful information sources related to a course, although many such sources exist in the rest of the world.

The progress in computer and information technology offers new perspectives for education [Romiszowski 1990, Reinhardt 1995]. By proper use of *multimedia*and *hypertext*, it is nowadays possible to present the educational material without the instructor's physical presence and, furthermore, in a way that is at the same time attractive to the students and pedagogically efficient. For example, a student having a personal computer with multimedia capabilities (appropriate graphics and sound cards) and a CD-ROM unit can watch recorded video of selected parts of lectures or experiments, listen to sound, run live simulations, and all these without ever going to amphitheatre or the laboratory [Speh et al. 1994]. However, many problems still remain. A student cannot submit questions, there is no collaboration nor competition with other students and generally the student is cut off from the rest of the educational team. In a few words, there is no sufficient interaction.

Recent developments of computer and information technology in the fields of *computer networks*and *networked hypermedia systems*promise to supply with attractive solutions to these problems. It is now possible to store the educational material in a central computer and allow it to be accessed by many users at the same time, through personal computers connected to a local- or wide-area network. Availability of the educational material is thus significantly improved and the material itself can be updated regularly in a much easier and economical way. The problem of interaction is challenged in various ways with the use of both asynchronous and synchronous communication [Hiltz 1995, Harasim et al. 1995, Capell 1995]. Learning environments implemented on those grounds shall be referred to as *Networked Learning Environments* (NLEs). Obviously, the computer network is of paramount importance in these environments as well as the networked hypermedia systems and the communications tools available on the network. The use of these NLEs renders feasible the offering of education to a much large number of interested students and, at the same time, reinforces the result of the instruction [Rossner et al. 1995]. They can be viewed as steps towards the ideal of *Open and Distance Learning* (ODL).

NLEs can be implemented in various different ways. One of them is by using existing components, building upon them and creating an integrated learning environment that provides the required functionality. An experimental environment of this kind is currently under development at the Software Engineering Laboratory of the Computer Science Division, at the National Technical University of Athens (NTUA). This particular environment is the subject of this paper and will be referred to as *Nov-NLE*. Nov-NLE will be employed in carrying out a specific project in ODL, named EONT [Papaspyrou et al. 1996], within the SOCRATES programme of the European Union.

2. Requirements and Design of Nov-NLE

The following requirements have been set for Nov-NLE:

- The educational material must be widely available and able to be accessed by many computers from many different locations at the same time. It must also be hierarchically structured and use hypertext links in such a way as to facilitate and guide the users. Furthermore, the environment must be equipped with efficient authoring tools to facilitate the development of courseware and must provide an easy way to manage and regularly update the educational material.

- The environment must provide the students with a means of submitting questions and the instructor with a means of answering them. It must implement a communications channel, in which all students and the instructor will participate and discuss matters related to a specific course electronically. The content of the discussions that take place, as well as all questions and answers, must be stored in such a way that they can be reused. The environment must also support collaborative work.

- The environment must facilitate the users in their attempts to locate and access course related material that is distributed in the Internet. It must also provide with a means of communicating with other scientists and students from the rest of the world.

- The environment must facilitate the automation of administrative tasks, specify access and update rights for its users and have a friendly user interface that will make things easy for the learner.

Nov-NLE was designed for flexibility. At a physical level, it is a distributed system and uses the client-server model [Sinha 1992]. There is a server, in which all material is centrally stored, and many clients, one for each user accessing this material. The clients need not be physically located in the Software Engineering Laboratory, but they can connect to it through the NTUA's network, through the Internet or via a regular phone line, as shown in [Fig. 1].

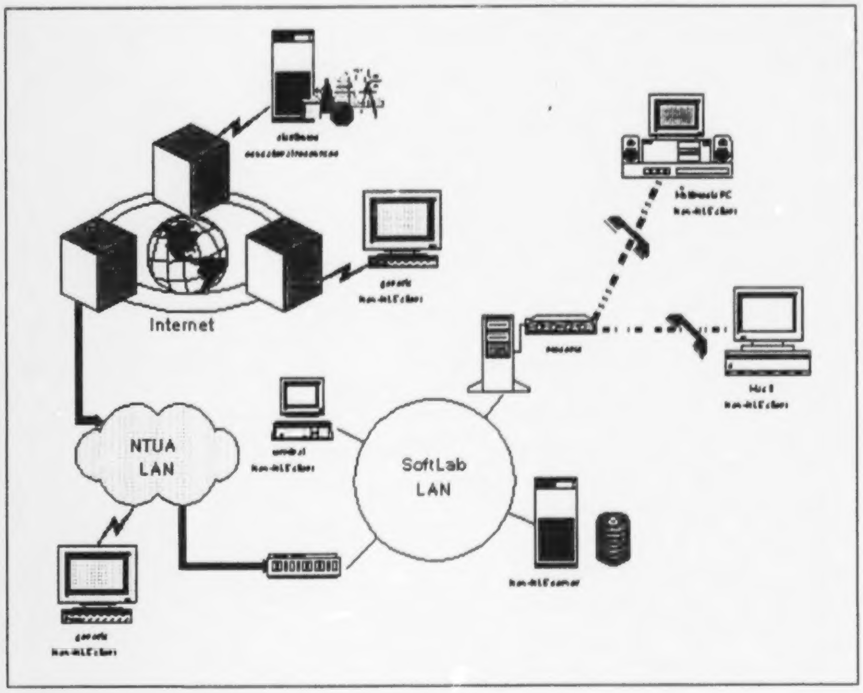


Figure 1: Client-server implementation at the physical level.

Special emphasis has been put on the structure of Nov-NLE's components, in such a way that each of the offered courses contains the educational material (electronic form of the book, lecture notes, experiments, simulations, case studies, etc.) and additionally provides access to related material that is distributed in the Internet. Nov-NLE consists of the following components, as shown in [Fig. 2]:

BEST COPY AVAILABLE

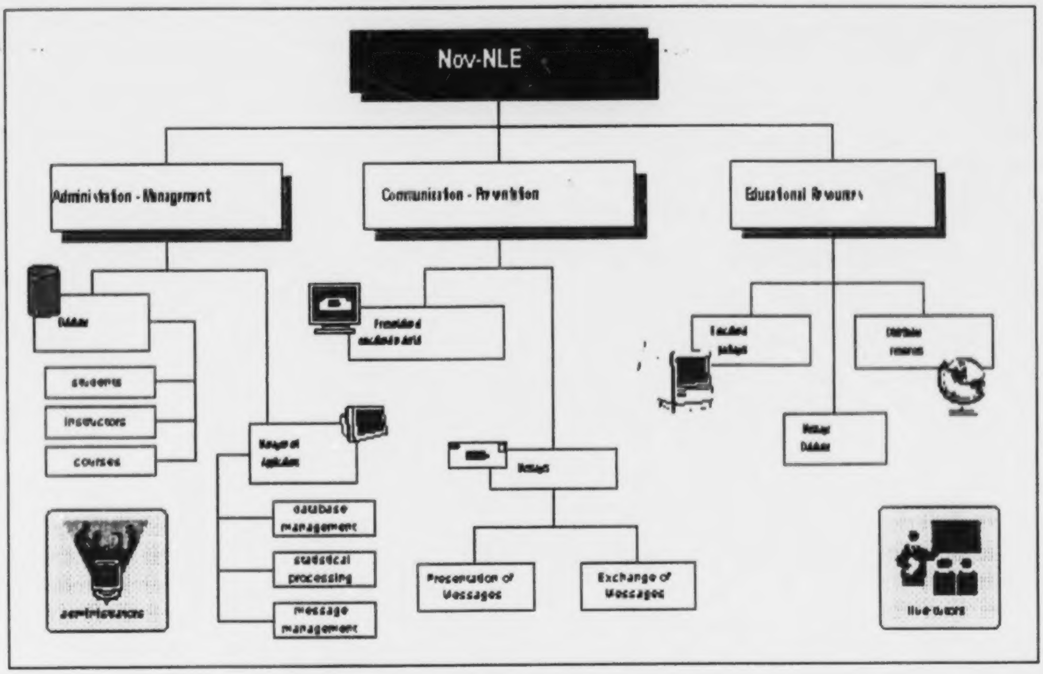


Figure 2: The structure of Nov-NLE.

- The *administration and management subsystem*, which supports the overall operation of Nov-NLE. Its main component is a database that holds all data necessary for the system's operation. Specifically, this database contains: (1) the *book of students*, which holds personal data for each student and detailed information about his/her progress in each course; (2) the *book of courses*, which holds information for each offered course; and (3) the *book of instructors*, which holds personal data for each instructor as well as other information (courses taught, supervised students, etc.). The role of this database in Nov-NLE is very important, since in addition it contains the users' access and update rights. Based on these rights it is possible to implement and operate a security control mechanism similar to that of a UNIX file system: the author of each object can specify what users are allowed to access and alter the object. The management of this database is performed by a set of management applications, to which only system administrators have access.

- The *communication and presentation subsystem*, which is responsible for the presentation of the educational material to the users of the system, as well as for the communication between the participants (instructor and students). Its major element is the man-machine interface. Since the design of Nov-NLE is based on the client-server model, this interface depends on the client that is used. Both text- and graphics-based clients are supported for a variety of platforms (IBM PC compatible under MS Windows, Macintosh, UNIX under X-Windows).

  The communication-presentation subsystem allows the exchange of various kinds of messages between Nov-NLE users. These messages offer a mechanism for communication between instructor and students and for submission of comments, questions, answers and remarks. Its proper use can partially cover the gap that is created by the lack of personal contact between students and the instructor [Harris 1995]. In order to make this message mechanism usable, three different types of messages are allowed:

  - *memos*: information that is automatically presented to a group of users as soon as they connect to the system. They are mainly used for public announcements by the instructors.
  - *private messages*: messages with a specific sender and recipient, which are not stored

BEST COPY AVAILABLE

permanently.

- *discussion messages*: messages of common interest, without a specific recipient, that can be read and answered to by all users. The system is responsible for storing such messages in an appropriate database, sorted by subject, in such a way that users can easily follow and participate in a discussion on a specific subject. Discussion messages are the core of the communication-presentation subsystem.

- The *educational resource subsystem*, which contains all the resources and relevant information used by Nov-NLE. It consists of three subsystems:
    - The *message database*, which stores all information about messages that are exchanged in the system. Users can seek information in comments, questions, answers and remarks that the instructor or other students have made.
    - The *educational package subsystem*, which contains educational packages for all offered courses. Each package contains the necessary educational material for a given course in hypermedia form. In addition, it can contain other information or software related to the course (bibliography, simulations, case studies, etc.).
    - The *distributed resources subsystem*, which contains references to educational resources that are distributed in the Internet. These resources can contribute to the instruction process, by means of "educational field-trips" in the Internet.

Finally, the human component of Nov-NLE consists of live tutors, responsible for the system's interactivity, and administrators.

3. A Prototype of Nov-NLE

A prototype of Nov-NLE is under construction at the Software Engineering Laboratory of NTUA. The prototype will be used for the purposes of the project EONT [Papaspyrou et al. 1996], in which participate seven universities from seven countries of the European Union. This prototype uses as infrastructure the Internet and the second-generation networked hypermedia system Hyper-G [Kappe et al. 1993]. The reason we chose Internet is because it offers a large number of services, it has a real treasure of information, and also is widely used and available today [Hesslop 1994]. The main reason we chose Hyper-G is because it possesses characteristics that facilitate the development of NLEs [Maurer 1996]. These characteristics are adequate for constructing an NLE that can meet all the requirements that were mentioned before. Specifically:

- It is a distributed system, in which the stored data can be placed at a variety of sites. There is no need for a central, dedicated server with huge amounts of disk storage.
- Hyperlinks are not stored within documents, but in a separate link database. This allows users to attach links to otherwise read-only documents (for example documents on a CD-ROM).
- Objects can be added to the system piecemeal as they are constructed.
- The system utilises access rights for every collection of hyper-documents.
- There are sophisticated search mechanisms.
- Any user of the system can annotate any document, either privately or publicly.
- The system is interoperable with other first generation hypermedia systems, such as WWW and Gopher, and with other popular Internet services, such as Telnet and FTP.
- The system is multilingual.

Software is also required for the clients of the prototype. For this purpose, existing software can be used. Possible options include the systems HM-Card, Harmony, Amadeus [Flohr 1995] or even the latest versions of popular WWW clients such as Netscape. In the first version of the prototype, whose front page is shown in [Fig. 3], the implemented components are the following:

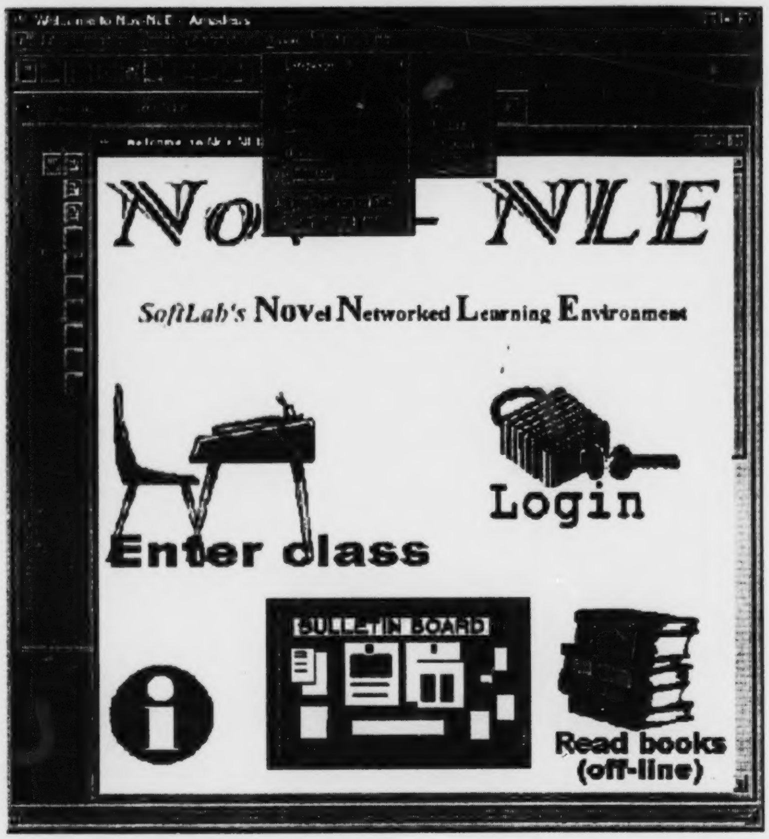


Figure 3: The front page of our prototype Nov-NLE (first version).

- *Database*: the current implementation of the database is based on a RDBMS.
- *Educational packages*: the current implementation contains one course package, with subject "Introduction to Software Engineering". However, for the purposes of EONT, each partner will develop one course package from the wide area of computer science.
- *Administration and management applications*: these applications are mainly used for the management of the database with a user-friendly graphical interface.

4. Concluding Remarks

There are a lot of things to improve concerning the state of education in our days. Some problems remain unsolved for many years and new ones arise, due to changing economic and social conditions. Higher education will profit a lot from the use of new technologies in computer networks and networked hypermedia systems. New learning environments can be built, particularly Networked Learning Environments which are the subject of this paper. The general and irreversible tendency of our age towards a future *Information Society* will help the development and use of such environments. In this Information Society, it is evident that information in digital form will require a lot of storage space and fast transfer rates. Considering the present state of information and network technology, this certainly can be thought of as a problem but its solution seems close since technology is rapidly advancing. The information highways that the European Union has committed to construct within the next 15 years [Commission 1994] are bound to help towards this direction, and the *Global Information Infrastructure* is presently a foreseeable goal. Nov-NLE is still an experimental system in an early stage of development and there are not adequate results from its use, in order to attempt an evaluation. It is certainly a step towards the implementation of usable learning environments, in the direction of the future Information Society. Environments like Nov-NLE are expected to play a significant role in the future and should be regarded today as forerunners of much more advanced environments that will soon emerge.

References

[Capell 1995]
Capell, P. (Sep. 1995) *Report on Distance Learning Technologies*. Technical Report CMU/SEI-95-TR-004. Software Engineering Institute, Carnegie Mellon University.

[Commission 1994]
Commission of the European Communities. (May 1994) *Europe and the Global Information Society: Recommendation to the European Council*, (also known as the Bangemann Report). Brussels.

[Ferguson 1975]
Ferguson, J. (1975) *The Open University from Within*. University of London Press Ltd.

[Flohr 1995]
Flohr, U. (Nov. 1995) Hyper-G Organizes the Web. *BYTE*. 59-64.

[Harasim et al. 1995]
Harasim, L., Hiltz, S. R., Teles, L. and Turoff, M. (1995) *Learning Networks: A Field Guide to Teaching and Learning Online*. MIT Press. Cambridge, MA.

[Harris 1995]
Harris, J. (1995) Organizing and Facilitating Telecollaborative Projects. *The Computing Teacher*, 22(5).

[Hesslop 1994]
Hesslop, B. (1994) *The Instant Internet Guide: Hands on Global Networking*. McGraw Hill. New York.

[Hiltz 1995]
Hiltz, S. R. (1995) *The Virtual Classroom: Learning Without Limits via Computer Networks*. Ablex Publishing Corporation. Norwood, NJ.

[Kappe et al. 1993]
Kappe, F., Maurer, H. and Scherbakov, N. (1993) Hyper-G: A Universal Hypermedia System. *Journal of Educational Multimedia and Hypermedia*, 2(1). 39-66.

[Maurer 1996]
Maurer, H. (1996) *Hyperwave: The Next Generation Web Solution*. Addison Wesley.

[Papaspyrou et al. 1996]
Papaspyrou, N., Koutoumanos, A., Maurer, H., and Skordalakis, E. (1996) An Ongoing Experiment in ODL Using New Technologies. In *Proceedings of the World Conference of The Web Society (WebNet'96)*.

[Reinhardt 1995]
Reinhardt, A. (Mar. 1995) New Ways to Learn. *BYTE*, 20(3). 50-72.

[Romiszowski 1990]
Romiszowski, A. J. (1990) Computer Mediated Communication and Hypertext: The Instructional Use of two Converging Technologies. *Interactive Learning International*, 6(1). 5-29.

[Rossner et al. 1995]

Rossner, V. and Egan, K. (1995) Using Computers in Teaching Sophisticated Conceptual Material via Distance Education, in *Proceedings of the Cambridge International Conference on Open and Distance Learning - Putting the student first: Learner-Centred Approaches in Open and Distance Learning*. 174-182. Cambridge.

[Sinha 1992]
Sinha, A. (Jul. 1992) Client-Server Computing. *Communications of the ACM*, 35(7). 77-98.

[Speh et al. 1994]
Speh, M., Carter, B., Reilly, C., and Wang, J. C. (1994) WWW for the Globewide Network Academy. In *Advance Proceedings of the First World-Wide-Web Conference*. 335-346. Geneva, Switzerland, May 25-27, 1994.

[Young 1995]
Young, M., (Feb. 1995) The Prospects for Open Learning. *Open Learning*, 10(1). Pitman Publishing.

Acknowledgements

The project EONT is partially funded by the European Commission's SOCRATES Programme. Its reference number is TM-OP-1995-1-GR-88 (1/0).

Embedded Interactive Concept Maps in Web Documents

Rob Kremer and Brian R. Gaines
Knowledge Science Institute
University of Calgary
Calgary, Alberta, Canada T2N-1N4

{kremer, gaines}@cpsc.ucalgary.ca

Abstract: Concept maps are an important means of knowledge representation. People find concept maps intuitive and easy to understand, and they are also amenable to formalization to provide computational services. Concept maps have been used in many fields including education, management, artificial intelligence, knowledge representation, knowledge acquisition, and linguistics. A concept map editor has been ported to the web as a Netscape plug-in where it provides embedded interactive maps within web documents. These maps may be used as an enhancement to web navigation, as a new live annotation form, and as a group interaction tool.

Introduction

Concept maps are an intuitive visual knowledge representation technique. Concept maps tend to be much easier for human users to understand than other knowledge representations such as pure text or predicate logic [Nosek 90]. Concept maps are graphs consisting of nodes with connecting arcs, representing relationships between nodes [Lamibotte 84]. The nodes are labeled with descriptive text, representing the "concept", and the arcs are labeled (sometimes only implicitly) with a relationship type.

Concept maps are used to structure argument forms and express relationships between ideas [Gaines 93]. In education, Novak and Gowin [Novak 84] have promoted the use of concept maps to investigate a student's understanding of a topic, and there are many different forms that have been applied in this field [Lambiotte 84]. In management, Axelrod [Axelrod 76] proposed *cognitive maps* as a means of representing the conceptual structures underlying decision making, and these have been used empirically to analyze organizational decision making [Eden 79] social systems [Banathy 91] and the policies of political leaders [Hart 77]. In artificial intelligence, Quillian [Quillian 68] developed a form of concept maps that came to be termed *semantic networks* and used extensively for formal knowledge representation.

In knowledge acquisition, concept maps are used to elicit knowledge from experts, for example the Wright-Patterson studies of the pilot's associate [McNeese 90]. In linguistics, Graesser and Clark [Graesser 85] have developed an analysis of argument forms in text in terms of structured concept maps with eight node types and four relation types. In the history of science, Thadgard [Thadgard 92] and Nersessian [Nersessian 89] have used the dynamics of concept maps to model processes of conceptual change in scientific revolutions. In the philosophy of science, Toulmin [Toulmin 58] developed a theory of scientific argument based on typed concept maps that provides a model of the rhetoric of Western thought [Golden 76].

Concept maps can represent knowledge on the very formal level, as in Gaines' KRS system, where concept map structures act as a complete interface to a knowledge representation language [Gaines 91] based on Classic [Borgida 89]. Another formal concept mapping system is the *GrIT* [Eklund 93] system that edits Conceptual Graphs [Sowa 84] and processes them using PEIRCE [Ellis 92]. But concept maps can also represent knowledge on a much less formal level if the system puts fewer constraints on the graphical "syntax". These two factors, intuitive understanding and both formal and informal knowledge representation, are of great importance: Both novices and experts can potentially use the same media. Novices are free to express themselves informally and without constraint, while experts can express themselves under the constraints of very formal semantics which allows for computational support.

Figure 1 shows a typical concept map from an educational domain in which a student has been asked to show the way in which they think about water [Novak 84]. The map has two types of nodes, *concepts* shown by ovals, and *instances* shown by rectangles. These are linked by arrows labeled with relations such as *needed by*, *made of*, *changes*, and so on. The conceptual structure developed encompasses some of the physics and biological roles of water. The student has developed the map within broad guidelines as to what are concepts and instances, and that they are to be linked by labeled directed arrows denoting relations.

Figure 1: Concept map of student's knowledge [Novak 84]

Concept Map Editing

In spite of the many virtues of concept mapping, and its use in so many fields, it is not ubiquitous. Why should this be so?

The claim has already been made that concept maps are intuitively easy for people to understand and manipulate. One would therefore conclude that concept maps should be common representational media today. But this not the case (at least not outside of specialized circles). The reason for this apparent contradiction is that, before the recent age of powerful graphical workstations, concept maps were difficult and time consuming to produce. Concept maps had to be carefully pre-planned to avoid excessive erasures or recopying when large swaths of the map had to be moved to make room for an unanticipated insertion. Even worse, revisions or extensions to concept maps were usually just as difficult to produce as the original because it usually involved recopying the entire map. Compounding the problem was the fact that large concept maps need to be broken up into a web of composite concept maps, and navigation among these inherently non-linear references was difficult and awkward when imposed on linear media such as books. These problems kept concept mapping techniques confined to static media such as printed media and specialized purposes such as student modeling in education and language analysis. Even these applications had to confine themselves to relatively small conceptual structures.

That was before the introduction of cheap and easily accessible computers with full color graphical interfaces. Graphical workstations remove the labor-intensive task of concept map design. Graphical concept map editing programs such as KDraw [Gaines 91], Kmap, WebMap [Gaines 95a, 95b], and Smart Ideas allow users to quickly slap down ideas (nodes and arcs) onto the workspace without having to carefully pre-plan the layout. These editors allow authors to easily slide concept map nodes (or groups of nodes) around the workspace: the editor takes care of maintaining relationships (arcs) among the nodes. These and other features, allow for easy creation, modification, and copying of concept maps by non-experts. Computer workstations can also be used as a display media for concept maps. This affords easy navigation between related concept maps and other media; user annotation and extension; and user feedback to the author.

Concept Mapping and Hypermedia: Smart Ideas

Smart Ideas [Kremer 93], a program which allows users to draw concept maps with a graphical direct-manipulation interface, has been applied to brainstorming, planning, knowledge elicitation, documentation, presentation, decision support, and meeting facilitation in such diverse areas as research planning, computer networking, marketing, education, oil and gas strategy planning, computer program design, formal language design, academic paper planning, literature analysis, and personal planning. The authors' experience to date has been that most people enthusiastically adapt to reading and understanding informal concept maps (which concurs with [Lambiotte 84]) and readily learn to create and modify concept maps. Kremer has introduced computer-supported concept mapping to individuals who could then independently draw concept maps within five minutes, and to groups using separate workstations with a shared concept map workspace who where collaborating on their topic area within fifteen minutes.

Smart Ideas allows one to draw concept maps using an editor, but every node and link drawn in the map is also represented in an underlying hyperspace which is potentially shared with other users and other concept maps. Figure 2 illustrates the concept. The hyperspace contains "abstract" hypernodes that have hyperlinks between them. In this case, the nodes represent states in a simple communication protocol. The links between them all represent state transitions. Surrounding the hyperspace are several concept maps as they would appear to a user reading or editing them. Each concept map shows some subset of the nodes and links in the hyperspace. The concept map window labeled "The Complete Protocol" contains a view of the entire hyperspace.

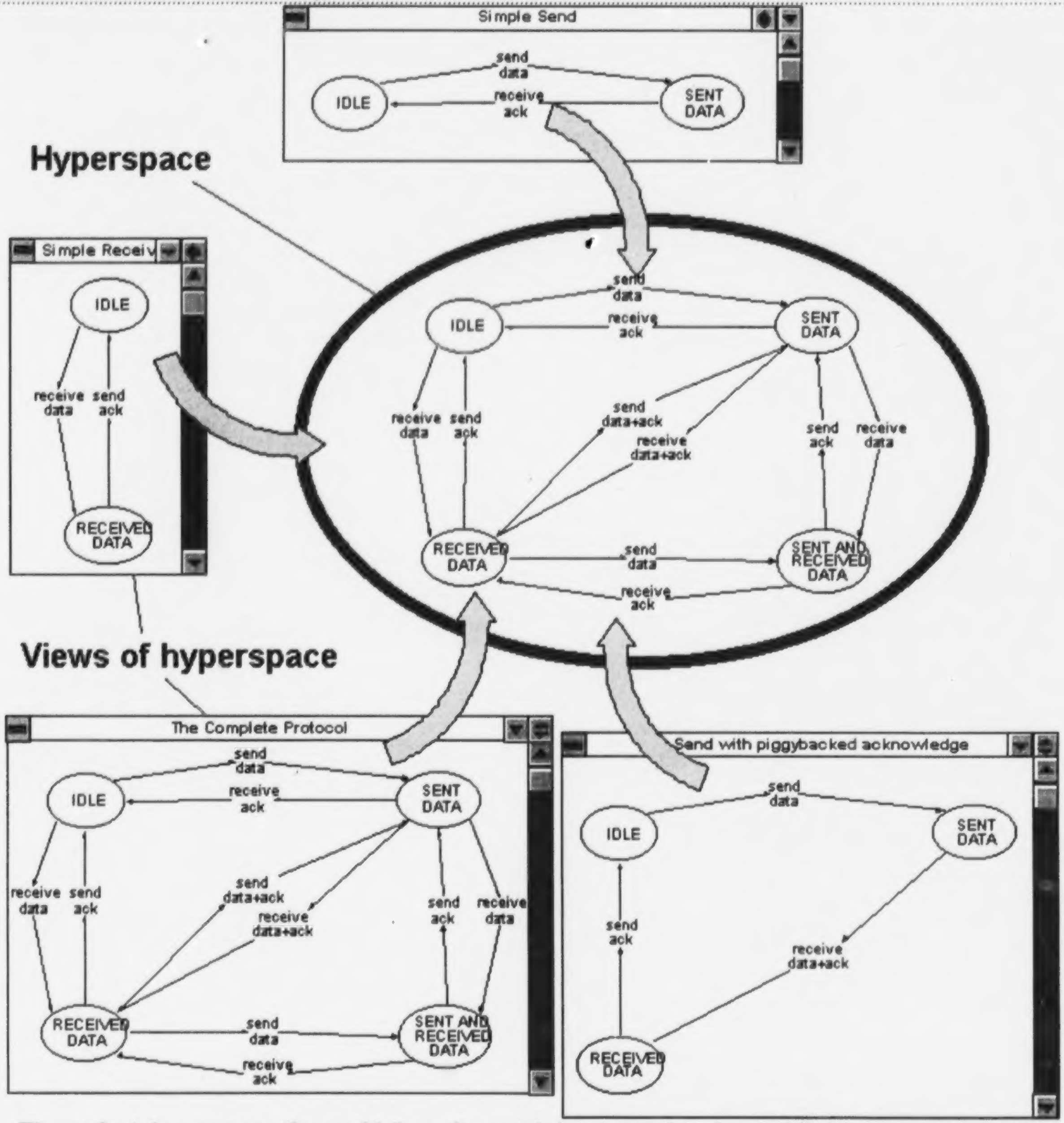

Simple Send
send data
receive ack
IDLE
SENT DATA
Hyperspace
Simple Receiv
IDLE
receive data
send ack
RECEIVED DATA
send data
receive ack
send data+ack
receive data+ack
send ack
receive data
SENT AND RECEIVED DATA
Views of hyperspace
The Complete Protocol
Send with piggybacked acknowledge

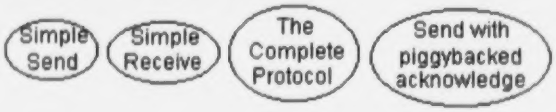
Figure 2: A hyperspace (large thick oval) containing several nodes and links represents a state transition diagram of a simple communication protocol. Four concepts are shown which contain subviews of the hyperspace.

Although Figure 2 shows similar graphical layouts between the concept maps, no such constraint is enforced by the system: users may drag nodes and link labels to achieve any layout they wish; the arcs always remain anchored to their terminating objects.

Whenever a user draws a new node in a concept map, a corresponding hypernode is always created in the hyperspace. Likewise, whenever a user draws a new link between two nodes in a concept map, a corresponding hyperlink is created between the corresponding hypernodes in the hyperspace. Nodes and links may be removed from the view with or without removing them from the hyperspace. Most operations on the concept maps, such as renaming nodes and links, have corresponding effects in the hyperspace.

One advantage of storing all the information in the hyperspace is that one can easily *expand* nodes in a concept map to reveal its neighbors in the hyperspace. An expansion operation on a concept map node looks up each of the hyperlinks that are anchored (either to or from) the corresponding hypernode. For each link, the other anchor node is either found in the current concept map or is drawn as a new concept map node, and the labeled link is drawn between the two anchor nodes. Users may request expansions to proceed recursively to any depth. For example, looking back to Figure 2, a user looking at the "Simple Send" view could perform a two-level expand of either node to have the system draw in all the information in the hyperspace (although the layout may not be optimal).

Smart Ideas uses the nodes in concept maps as a navigational tool. One can merely double-click on a node to hyperlink to an associated document. The document is typically the Smart Ideas hypernode which is represented by the concept map node. This may contain another concept map. Indeed, this answers the question of where Smart Ideas concept maps are stored-they are always stored as the contents of hypernodes in the hyperspace. A hyperspace contains nodes which contain views of the hyperspace itself. It might be pointed out at this time that the information in Figure 2 is a bit rarefied: assuming all of Figure 2's views reside in the same hyperspace, the hyperspace is a bit bigger than advertised, for it must also contain one node for each view:



In addition to creating drawing concepts, Smart Ideas allows users to convey information using text or arbitrary multimedia using the OLE protocol. These multimedia objects may be used either stand-alone or as annotations to concept maps.

Smart Ideas on the Web

Smart Ideas has recently been ported to the web as a Netscape plug-in application. Thus, the concept maps can not only be viewed, but also manipulated and used to navigate the Smart Ideas hypermedia database. Although, at the time of writing, the Smart Ideas plug-in is a read-only version of the Smart Ideas hypermedia, there are many operations available to the user. These include:

- rearranging the map by dragging nodes and link handles,
- double-clicking a node to navigate to deeper detail about a particular concept,
- querying a node about its relationships to other nodes to display all links anchored on that node as well as the related nodes.

Like all plug-in application, a Smart Ideas plug-in can be part of a HTML document (see figure 4) via the <embed> HTML directive, or it can be viewed as a document itself, filling the entire client area of the Netscape browser. To embed a Smart Ideas document in an HTML document one uses the following HTML code:

```
<embed src="test.sbk" width=500 height=150>
```

where "test.sbk" is the URL or a Smart Ideas "book" (file) and the width and height values can be any appropriate value in pixels. To view a Smart Ideas document directly one needs only navigate to the appropriate URL with the "application/smartIdeasBook" MIME type.

Every Smart Ideas document has a designated "home page", and this page is opened by default whenever the document is first displayed. From there, one can navigate through the rest of the Smart Ideas hypertext. Navigation is accomplished by simply double-clicking on a node, which causes the Smart Ideas display (whether embedded in an HTML document or as a full size document) to be replaced with the new page referenced by the concept map node. Note that, in the usual case of the node referencing a page within the same Smart Ideas document, this page change does not force a URL access, but is strictly local within the Smart Ideas sub-window: Netscape's back button will not cause the Smart Ideas page to roll back, and enclosing document context is strictly maintained. Since a user may none-the-less wish to roll back Smart Ideas pages, a back button is proved at the bottom of every Smart Ideas window (as part of a caption bar that always displays the name of the current page).

Concept map node references may reference a page within its document, a page in another Smart Ideas document, or any external URL. Unlike the first case, the later two cases will cause a URL fetch to be generated and Netscape will handle the paging as it does with any normal hyperlink event.

This navigation design was chosen to allow Smart Ideas concept maps to be used flexibly as either a web navigational aid (using external node references) or as a "live picture" concept map within a larger context (using internal node references). As an example of concept maps as web navigational aids, Figure 3 is a two-level breakdown of the documents navigable from the "Plug-in SDK" page on Netscape's web site. This concept map gives a quick and clear picture of the structure of the document set and may allow a user to directly navigate to the page of interest without having to traverse, scan, and scroll through all the intervening documents. The node types (as reflected by color and shape - HTML documents are yellow ellipses, GIF files are white rectangles, etc.) also assists the user to zero in on the appropriate document.

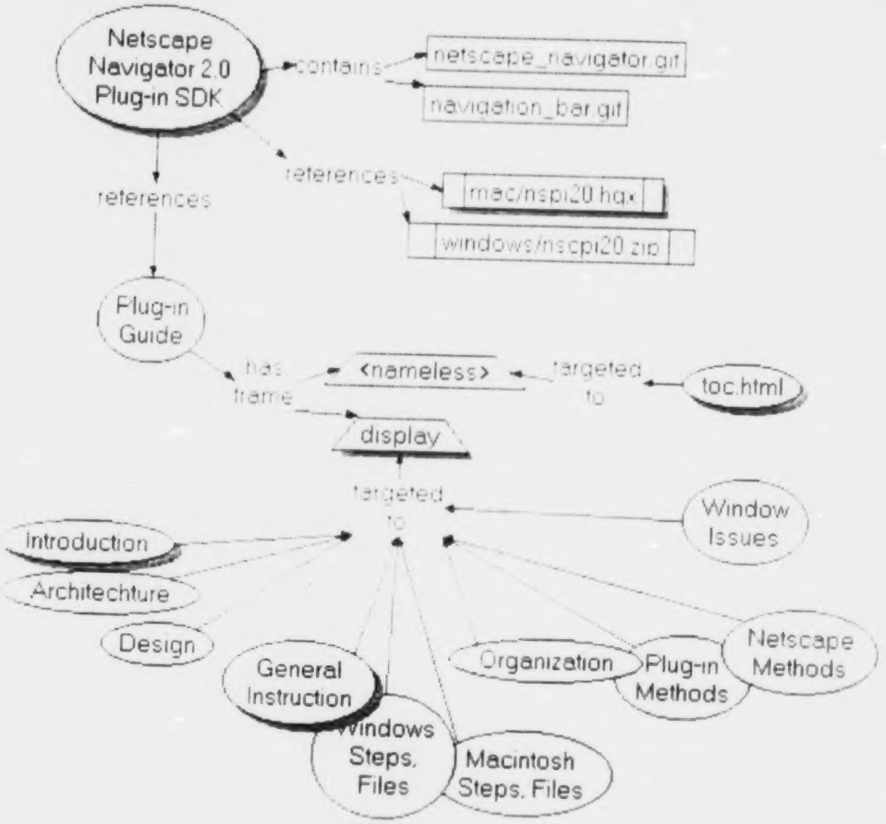
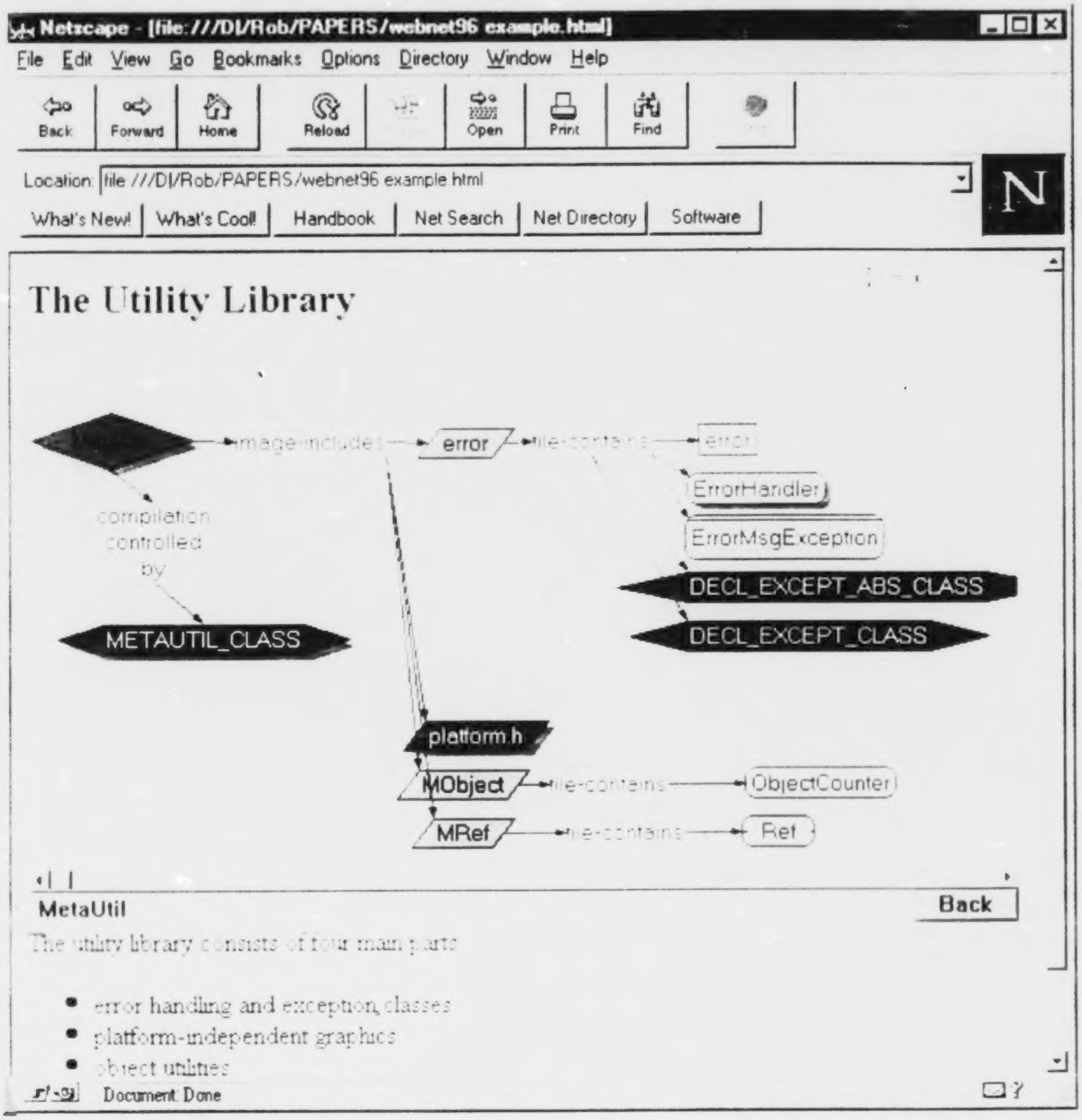

Netscape Navigator 2.0 Plug-in SDK
contains
netscape_navigator.gif
navigation_bar.gif
references
mac/nspi20.hqx
windows/nscpi20.zip
references
Plug-in Guide
has frame
<nameless>
targeted to
toc.html
display
targeted to
Window Issues
Introduction
Architechture
Design
General Instruction
Windows Steps, Files
Macintosh Steps, Files
Organization
Plug-in Methods
Netscape Methods

Figure 3: A two-level breakdown of the documents navigable from the Plug-in SDK page on Netscape's web site.

Figure 4 is an example of a live concept map embedded in an HTML document. The HTML document is a description of set of C++ class libraries. The concept map is embedded as a graphic picture of one of the libraries, but users can obtain more detail about any of the concepts in the map by double-clicking on them to replace the overview picture without loosing their context in the enclosing HTML document. The sub-window back button can be used to navigate back to the original view.

BEST COPY AVAILABLE


Netscape - [file:///D|/Rob/PAPERS/webnet96 example.html]
File Edit View Go Bookmarks Options Directory Window Help
Back Forward Home Reload Open Print Find
Location: file:///D|/Rob/PAPERS/webnet96 example.html
What's New! What's Cool! Handbook Net Search Net Directory Software
N
image-includes
error
file-contains
error
ErrorHandler
ErrorMsgException
DECL_EXCEPT_ABS_CLASS
DECL_EXCEPT_CLASS
compilation controlled by
METAUTIL_CLASS
platform.h
MObject
file-contains
ObjectCounter
MRef
file-contains
Ref
Back
Document: Done

The Utility Library

MetaUtil

The utility library consists of four main parts

- error handling and exception classes
- platform-independent graphics
- object utilities

Figure 4: An example of a live concept map embedded in an HTML document

Future Work

Live concept maps on the web suggest several areas of future work.

The web map shown in Figure 3 suggests a tool could be build to automatically generate a concept map given an HTML document as a starting point. The tool would examine the HTML document for contained objects and references to other pages, entering these into the Smart Ideas hypertext and views. Either these maps could be generated and used on-the-fly, or the generated Smart Ideas document could then be edited (using the stand-alone version of Smart Ideas) and used an index into the HTML document set.

BEST COPY AVAILABLE

The stand-alone version of Smart Ideas encourages users to not only read and navigate concept maps, but also to extend and annotate them. There is no reason why sufficiently privileged users could not do the same using the web version of the program. This would be an excellent tool for distance education and other forms of group interaction.

Once one starts to think of web concept maps as editable, one cannot help but jump to the vision of groupware concept mapping tool on the web. Such a tool might allow more than one user full interactive access to a map with WYSIWIS (what you see is what I see) interaction: Every time one user moved or otherwise edited a part of a concept map, every other user who happened to be viewing the same map would have the update automatically animated on their view of the map. This could be accomplished through the use of special registration servers that the plug-in would automatically query for a list of other current users of the map. Socket connection would then be established between the plug-ins.

Another area of research currently under way is the construction of graph theory which would allow users to constrain the extension of a graph such that a single tool could emulate many of the formal graph theories such as KRS [Gaines 91] and Conceptual Graphs [Sowa 84]. Blending this work with concept maps on the web would be powerful tool which would allow users to develop ideas informally, and then cast these ideas in an appropriate formalism for automated support and development [Kremer 94, 95].

One drawback of the current web version of Smart Ideas it has been ported only to Microsoft Windows environment. Ports to other platforms would be desirable, and one very good way to accomplish this may be to port the code from C++ to Java [Sun 95]. While Java has a few drawbacks (such as slower speed) its portability and automatic loading over the web offers significant advantages over custom C++ code. Porting to Java will be investigated in the near future.

Conclusion

Concept maps are a valuable and important means of knowledge representation. They have significant potential as an enhancement to web navigation, as a new live annotation form, and as a group interaction tool.

References

[Axelrod 76] R. Axelrod, *Structure of Decision*, Princeton, New Jersey: Princeton University Press. 1976.

[Banathy 91] B.H. Banathy, "Cognitive mapping of educational systems for future generations," *World Futures*, vol. 30, no. 1 pp. 5-17, 1991.

[Borgida 89] A. Borgida, R. J. Brachman, D. L. McGuiness, L. A. Resnick. "CLASSIC: A Structural Data Model for Objects." *Proceeding of 1989 SIGMOD Conference on the Management of Data*, New York, ACM Press, pp. 58-67.

[Eden 79] C. Eden, S. Jones, and D. Sims, *Thinking in Organizations*, London: Macmillan. 1979.

[Ellis 92] G. Ellis, R. Levinson, eds. *Proceedings of the First International Workshop on PEIRCE: A Conceptual Graphs Workbench*, Las Cruces, New Mexico, 1992.

[Eklund 93] P. W. Eklund, J. Leane, C. Nowak. "GRIT: Toward a Standard GUI for Conceptual Structures." *Proceedings of the Second International Workshop on PEIRCE: A Conceptual Graphs Workbench*, Laval University, Quebec, Canada, August 7, 1993.

[Gaines 91] B. R. Gaines. "An Interactive Visual Language for Term Subsumption Languages."

Proceedings of IJCAI-91, Sydney, Australia, August 24-30, 1991.

[Gaines 93] B. R. Gaines and M. L. G. Shaw, "Supporting the creativity cycle through visual languages," in *AAAI Spring Symposium: AI and Creativity*. AAAI: Menlo Park, California. p. 155-162, 1993.

[Gaines 95a] B. R. Gaines and M. L. G. Shaw. Concept maps as hypermedia components. International Journal Human-Computer Studies, 43(3), 323-361. 1995.

[Gaines 95b] B. R. Gaines and M. L. G. Shaw. WebMap: concept mapping on the web. World Wide Web Journal 1(1) 171-183. 1995.

[Golden 76] J.L. Golden, G.F. Berquist, and W.E. Coleman, *The Rhethoric of Western Thought*, Dubuque, Idaho: Kendall/Hunt. 1976.

[Graesser 85] A.C. Graesser and L.F. Clark, *Structures and Procedures of Implicit Knowledge*, New Jersey: Ablex. 1985.

[Kremer 93] R. Kremer. "A Concept Map Based Approach to the Shared Workspace." MSc. Thesis, University of Calgary, Canada, June, 1993.

[Kremer 94] R. Kremer. "Concept Mapping: Informal to Formal." *Proceedings of the Third International Conference on Conceptual Structures, Knowledge Represetnation Workshop,* University of Maryland, 1994.

[Kremer 95] R. Kremer. "The Design of a Concept Mapping Environment for Knowledge Acquisition and Knowledge Representation." *Proceedings of the Banff Knowledge Acquisition Workshop,* Banff, Alberta, Canada. 1995.

[Lambiotte 84] J. G. Lambiotte, D. F. Dansereau, D. R. Cross, S. B. Reynolds. "Multirelational Semantic Maps." *Educational Psychology Review,* **1**(4), 1984, pp. 331-367.

[Hart 77] J.A. Hart, "Cognitive maps of three latin american policy makers," *World Politics*, vol. 30, no. 1 pp. 115-140, 1977.

[McNeese 90] M.D. McNeese, B.S. Zaff, K.J. Peio, D.E. Snyder, J.C. Duncan, and M.R. McFarren, An Advanced Knowledge and Design Acquisition Methodology for the Pilot's Associate. Harry G Armstrong Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio. 1990.

[Nersessian 89] N.J. Nersessian, "Conceptual change in science and in science education," *Synthese*, vol. 80, no. 1 pp. 163-184, 1989.

[Nosek 90] J.T. Nosek, I. Roth. "A Comparison of Formal Knowledge Representation Schemes as Communication Tools: Predicate Logic vs. Semantic Network." *International Journal of Man-Machine Studies*, **33,** 1990. pp. 227-239.

[Novak 84] J.D. Novak and D.B. Gowin, *Learning How To Learn*, New York: Cambridge University Press. 1984.

[Quillian 68] M.R. Quillian, "Semantic memory," in *Semantic Information Processing,* M. Minsky, Editor. MIT Press: Cambridge, Massachusetts. p. 216-270, 1968.

[Sowa 84] J. F. Sowa. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, Reading, Massachusetts, 1984.

[Sun 95] Sun Microsystems Inc. http://java.sun.com/1.0alpha3/doc/overview/java/index.html. 1995.

[Thadgard 92] P. Thadgard, *Conceptual Revolutions*, Princeton, New Jersey: Princeton University Press. 1992.

[Toulmin 58] S. Toulmin, *The Uses of Argument*, Cambridge, UK: Cambridge University Press. 1958.

Building Internet-Based Electronic Performance Support for Teaching and Learning

James M. Laffey
304 Townsend Hall
University of Missouri-Columbia
Columbia, MO 65211
email: ciLaffey@showme.missouri.edu

Dale Musser
350 Townsend Hall
University of Missouri-Columbia
Columbia, MO 65211
email: DRMUSSER@aol.com

Abstract: The College of Education, University of Missouri-Columbia is developing and testing a suite of tools that utilize the Internet and work as a system to support learning from field experiences. These tools are built to support pre-service teachers, field-based mentors and college faculty as they collaborate, engage in practice, document their efforts, share their experiences, and assess outcomes. The journal system enables the preservice teacher to record their observations and reflections about experiences, maintain these records on a central server, and organize the records to meet a variety of needs. Among the resource types are web pages, electronic messaging and conferencing. The field test version of the journal system will implement a Macintosh client and Silicon Graphics Indy servers. The clients and servers communicate over the internet using TCP/IP connections.

Introduction

Educational and training organizations are moving away from sole reliance upon classroom teaching models and toward supporting learning in the context of authentic practice and work experiences. Technology is viewed as scaffolding for performance and an enabler of communication and feedback for learning. Gloria Gery [Gery 1991] coined the term electronic performance support systems (EPSS) for systems which provide on-demand access to integrated information, guidance, advice, assistance, training, and tools to enable high level job performance with a minimum of support from other people. For Gery the goal of an EPSS is to provide whatever is necessary to generate successful performance and learning at the moment of need. Critical to this work is to design the resources to fit the work.

Organizations justify EPSS as a means for building and insuring worker competence in the face of new competitive environments with increasingly difficult job demands, rapidly changing knowledge bases, and inadequate training systems. Inherent in this approach is the recognition that simply improving training or doing more training is not adequate. Organizations need a new approach to building and supporting competency in the work place and have been turning to EPSS. We see many parallels between the needs of the knowledge workers in the organizations that have been adopting EPSS and the needs of the teacher as they support new forms of learning in educational environments. Teachers and learners in educational settings need opportunities to put knowledge and skills to work in authentic, field-based projects. Success in these situations calls for teachers and students being resourceful in ways never imagined in the past and also being supported as they make sense of feedback and reflect on their experiences. The feedback from efforts and the reflection process are indeed the most significant products of the educational product.

Based on research and development undertaken at Apple Computer, support systems can augment the processes of taking on new challenges, sharing resources and being guided to successful performance.

[Laffey, 1995] articulated the characteristics of a dynamic support system as the ability to change with experience, to be updated and adjusted by the performer, and to augment other resources found in the performer's community. New models for education require that the support system also facilitate reflection.

An Internet-Based Support System

We envision Internet-based tools as a new form of EPSS, creating new performance and learning spaces which augment communities of practice. The Internet provides the potential of wide access and links to many types of dynamic and evolving resources. The Internet, however, has some limitations for supporting EPSS. Internet-based tools typically allow a community to create an accessible information space which is a starting point for our design of a performance and learning space. By information space we generally mean a place for the storage and retrieval of information -- to support information requests. By performance and learning space we mean a place for the storage, retrieval, *processing* and *reflection* upon information and experience -- to support requests which come in the context of planning a task, solving a problem, making a decision, trying a new technique or strategy, or assessing outcomes. In some instances the difference may be as subtle as the quality achieved by user-centered design for demanding users [Sellers, 1994]. In other instances it may be illustrated by new types of software designed for specific processes encountered during the performance [Gallopoulos, Houstis, Rice, 1994]. To this end the performance and learning space for supporting teaching and learning in authentic, field-based projects should include tools for managing and monitoring the complexity of projects (goals, milestones, time frames, resources, multimedia artifacts, etc.) and for building the students' skills and capacity for taking on and being successful at projects.

We are developing an interactive journal system for supporting students, field-based mentors and college faculty as they collaborate, engage in practice, document their efforts, share their experiences, and assess outcomes. The first version of the journal software enables the student-teacher to record observations and reflections about experiences, maintain these records on a server, and organize the records to meet a variety of needs. The software also facilitates sharing these records with faculty, mentor teachers and other student-teachers. In addition the software provides access to a variety of resources for enhancing experiences and solving problems. Among the resource types are on-line archives of knowledge about teaching and learning, links to other appropriate on-line archives, references to off-line support material, and electronic messaging and conferencing. A key aspect of the design and development work will be to implement support for encouraging and improving students processes of reflection upon their experiences.

Software Description

The first version of the journal system will implement a Macintosh client and Silicon Graphics Indy servers. The clients and servers will communicate over the internet using TCP/IP connections. Anyone who has access to the internet via a direct connection or a SLIP or PPP connection will be able to participate. In addition the software will support the creation and editing of journal entries off-line for later upload when a connection is available.

In addition to the features and functions that are supplied by the journal system client connection with the journal system server, the software integrates both web-browsing and e-mail for a full-functioned environment that supports communication and collaboration with the central mechanism being the shared journal. The following categories of functions and features are being implemented in the journal system:

- the creation and editing of multimedia journal entries. In the first version text, images, and web links will be supported. In the following version audio and video will be supported.

- the sharing of journal entries with other members of the community

- the ability to add comments and feedback to a journal entry

- the designation of tasks that a member of the community is to complete. For example, a mentor at the university may ask a pre-service teacher to make observations of a class and write pre-observation, observation, and post-observation journal entries

- access to informative and instructional resources: web pages, documents, and software

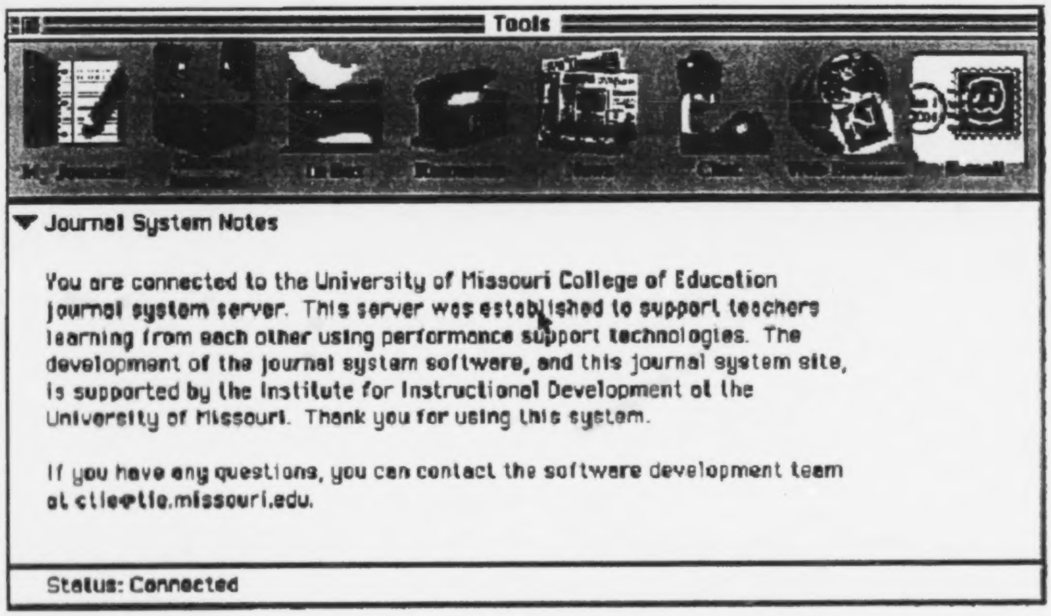
- access to e-mail within the journal client

- access to a web-browser within the journal client

- a multi-way text chat facility for real-time interaction from the journal client

- access to news and information that is important to the community of users

Figure 1 shows the basic tools available to students once they have signed onto the system. The tools include access to their own journal entries, access to journal entries from others and an In Box for notification about comments from others or newly assigned tasks. The other tools allow easy access to internet-based resources and communication tools.



You are connected to the University of Missouri College of Education journal system server. This server was established to support teachers learning from each other using performance support technologies. The development of the journal system software, and this journal system site, is supported by the Institute for Instructional Development at the University of Missouri. Thank you for using this system.

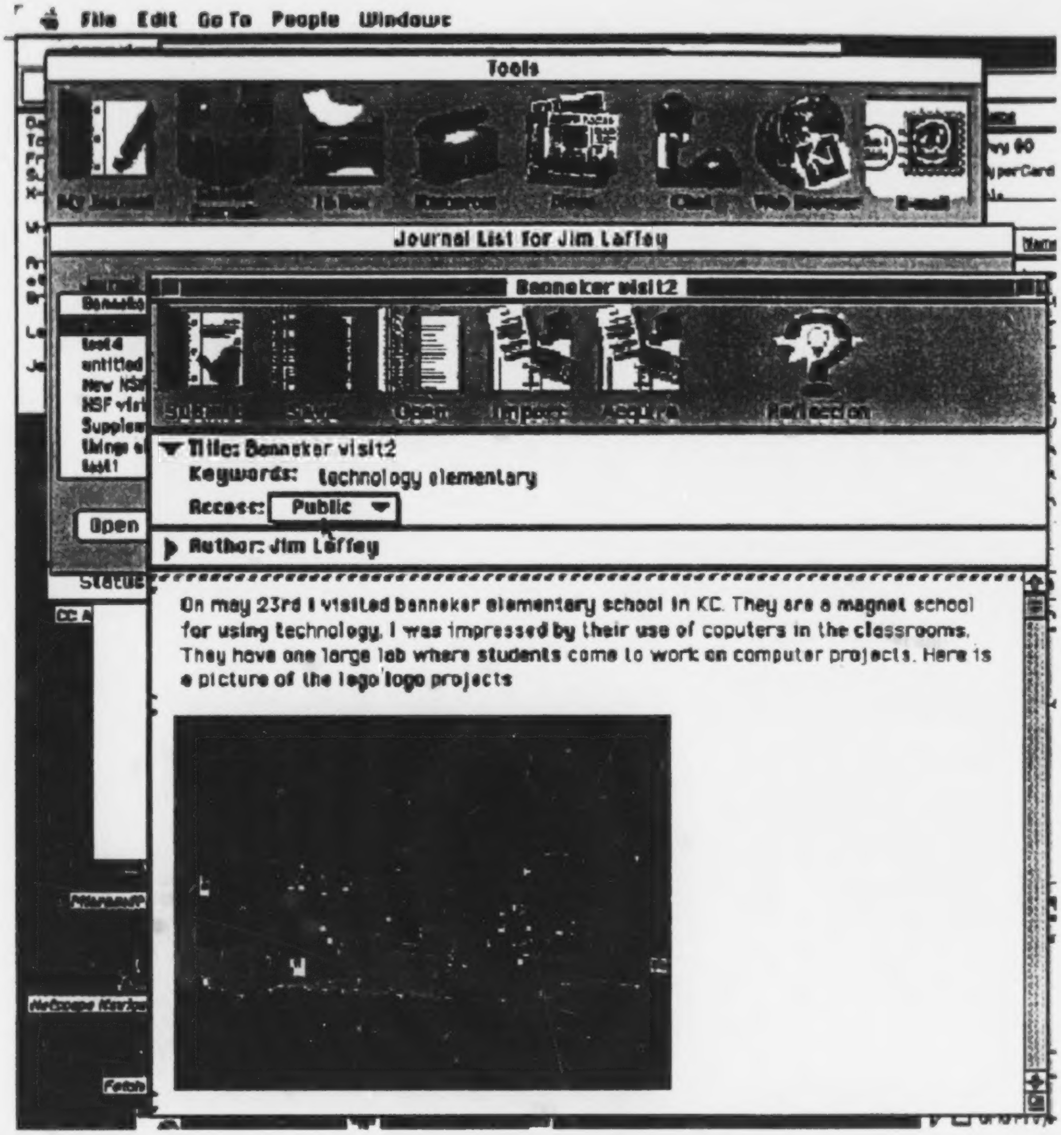
If you have any questions, you can contact the software development team at ctle@tle.missouri.edu.

Status: Connected

Figure 1: Opening screen

The journal entry screen (Fig. 2) enables the student to create a journal entry which can include imported graphics and links to URLs. The student can also view notes appended by the teacher and other students to the journal entry. The journal editor is designed to support and promote reflection; reflection is a key attribute of the journal writing activity. The reflection button takes evokes suggestions and mechanisms for reviewing your entry and thinking "deeper" about its meaning and significance. The editor allows the student teacher to import and acquire "live" media elements. The editor also support working in a networked environment for submitting entries to the server or working off-net with the capability of saving work to a hard drive for sending to the server when a connection can be made. The entry form allows the student to identify keywords to support search strategies and to set access rights to be public, private or to select groups.

BEST COPY AVAILABLE



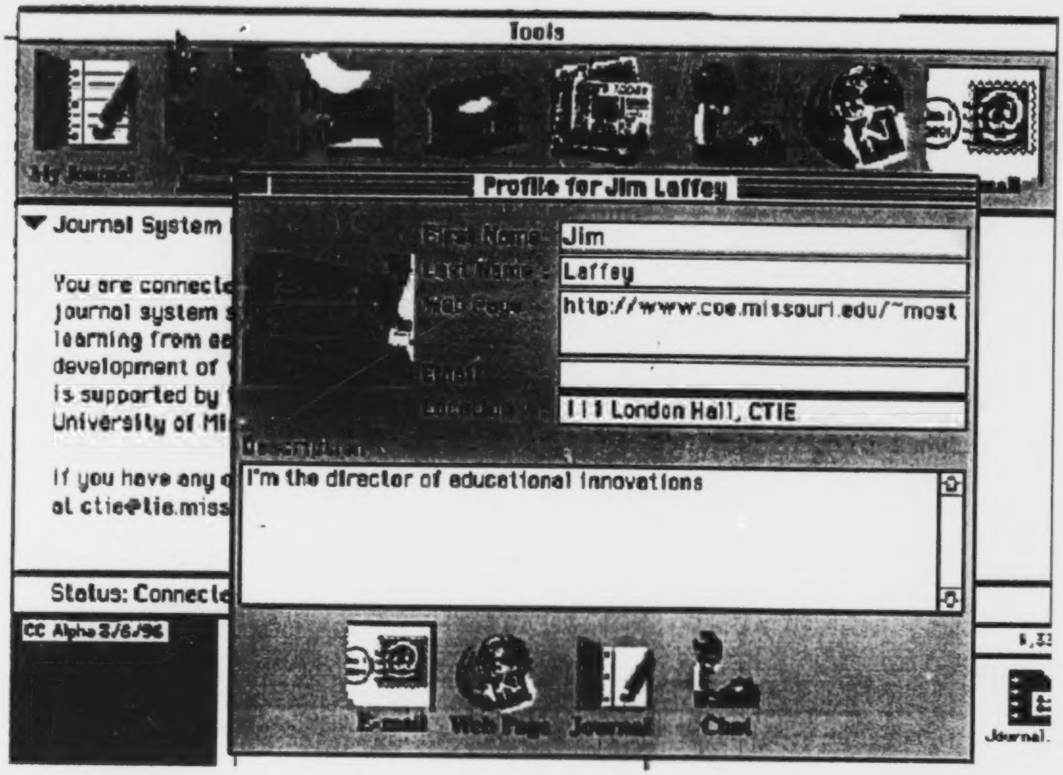
Title: Banneker visit2
Keywords: technology elementary
Access: Public
Author: Jim Laffey

On may 23rd I visited banneker elementary school in KC. They are a magnet school for using technology. I was impressed by their use of coputers in the classrooms. They have one large lab where students come to work on computer projects. Here is a picture of the lego logo projects

Figure 2: Journal entry

The journal system enables a shared set of documents for collaboration and reflection. the system is also intended to facilitate a shared community of educational practitioners for shared professional development. Much of the communications infrastructure is intended to support this goal. The journal system integrates the communication tools. Figure 3 which shows a profile of an author illustrates this integration by showing how from this profile one can access the author by using a preaddressed email, locating the authors home page, entering the authors journal or (if the author is on-line) connecting for a chat session. These communication functions enable a dynamic, interactive environment for learning and development.

BEST COPY AVAILABLE



You are connecte
journal system s
learning from ea
development of
is supported by
University of Mi

If you have any q
at ctie@tie.miss

Status: Connecte

CC Alpha 3/6/96

First Name: Jim
Last Name: Laffey
Web Page: http://www.coe.missouri.edu/~most
Email:
111 London Hall, CTIE
Description:
I'm the director of educational innovations

Figure 3: Author profile

Conclusions

The shared journal systems provides an opportunity for new forms of teaching and learning, forms which provide support for learning by doing and for working in the performance environment (typically outside the college classroom). These new forms of teaching and learning have much in common with electronic performance support systems which are now being implemented in business and industry. Technology can be an aid to improved human competency and performance by providing support for articulation, communication, processing, analysis, and reflection on human experience. The ISJS implements this type of support structure and offers great potential for bringing dynamic support systems to bear on the problems of aiding and improving human performance.

References

[Gery 1991] Gery, G. (1991). Electronic Performance Support. Cambridge, MA: Ziff Institute.

[Laffey, 1995] Laffey, J. M. (1995). Dynamism in performance support systems. Performance Improvement Quarterly, 8(1), 31-46.

[Gallopoulos, Houstis, Rice, 1994] Gallopoulos, E., Houstis, E., and Rice, J. (1994). Computer as thinker/doer: Problem-solving environments for computational science. Computational Science & Engineering, 1.2, 11-23.

[Sellers, 1994] Sellers, M. (1994). Designing for demanding users. Interactions, 1.3, 55-64.

BEST COPY AVAILABLE

Concurrent Manufacturing on the Web

Andrew Z. Lapsley, Brian R. Gaines and Douglas H. Norrie

Knowledge Science Institute and Division of Manufacturing Engineering
University of Calgary
Alberta, Canada T2N 1N4
drewl@insource.com, gaines@cpsc.ucalgary.ca, norrie@enme.ucalgary.ca
http://ksi.cpsc.ucalgary.ca/KSI/

Abstract: This article presents a system for the concurrent support of the manufacturing life cycle in a distributed enterprise. The World Wide Web is used as a knowledge transport mechanism to access distributed sites and data files, and the Netscape browser application is used as a file and data serving agent on the Web. A client helper has been implemented to provide a visual language interface to the Web, called XConMap, which allows users to dynamically link concept nodes in a concept map editor to Internet sites and files. In supporting and linking the data sets presented, the system is shown to be flexible and extensible through its open architecture, and is demonstrated working with existing application software and protocols.

- 1 Introduction
- 2 Mediator Architecture
- 3 Managing the Manufacturing Life Cycle with Mediator
- 5 Conclusions
- References

1 Introduction

Concurrent engineering (CE) is a philosophy and methodology of considerable interest in the manufacturing engineering community (Kusiak, 1993). Concurrent engineering, also referred to as simultaneous or parallel engineering, is a process which attempts to incorporate various life cycle decisions in the design of a product from the earliest stages onward. These stages include, but are not limited to, manufacturing, assembly, and maintenance. Knowledge based methods are increasingly used in order to ensure that all design changes and implementations are cohesive and complementary, and can be implemented with minimal effect on other stages of design and manufacture.

As the modern industrial environment moves towards a *post* mass-production paradigm (Tomiyama, 1992) issues like product maintainability and manufacturability become increasingly important. This importance is mainly in terms of achieving high levels of efficiency and marketability. Product cycle times and development times must be shortened and these activities made more efficient. In order to achieve the design goals of the concurrent engineering philosophy, several disparate groups and systems must be able to work together on a product, sometimes with competing goals, and almost always with highly diverse backgrounds. If concurrent engineering practices are to succeed, methods of knowledge representation and knowledge sharing must be applied to effectively utilize these human, technical and informational resources.

One approach to managing the information load and communication issues in a concurrent engineering endeavor is Mediator (Gaines, Norrie and Lapsley, 1995), developed as in the GNOSIS Consortium in the international Intelligent Manufacturing Systems (IMS) research program (Gaines and Norrie, 1995). Mediator is an open architecture information and knowledge management system designed to provide a flexible technology to support the management of complex distributed activities. It has been applied to supporting manufacturing activities throughout the product life cycle (Gaines et al., 1995), and to distributed education through the learning web (Norrie and Gaines, 1995). Mediator was originally implemented using custom software on local area networks. This article describes Mediator1, an implementation on the World Wide Web supporting international collaboration in distributed concurrent manufacturing.

2 Mediator Architecture

Figure 1 shows the original Mediator architecture (Gaines and Norrie, 1994) on the left and the web implementation on the right.

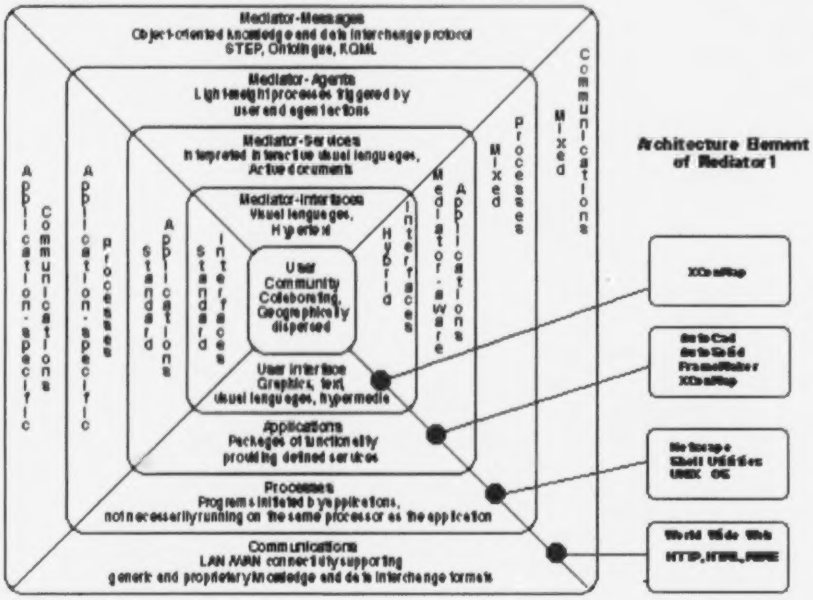

Mediator-Messages
Object-oriented knowledge and data interchange protocol
STEP, Ontolingua, KQML
Mediator-Agents
Lightweight processes triggered by
user and agent actions
Mediator-Services
Interpreted interactive visual languages,
Active documents
Mediator-Interfaces
Visual languages,
Hypertext
User
Community
Collaborating,
Geographically
dispersed
User interface
Graphics, text,
visual languages, hypermedia
Applications
Packages of functionality
providing defined services
Processes
Programs initiated by applications,
not necessarily running on the same processor as the application
Communications
LAN/WAN connectivity supporting
generic and proprietary knowledge and data interchange formats
Application-specific
Communications
Application-specific
Processes
Standard
Applications
Standard
Interfaces
Hybrid
Interfaces
Mediator-aware
Applications
Mixed
Processes
Mixed
Communications
Architecture Element
of Mediator1
XConMap
AutoCad
AutoSolid
FrameMaker
XConMap
Netscape
Shell Utilities
UNIX OS
World Wide Web
HTTP, HTML, MIME

Figure 1 Mediator implementation on the World Wide Web

The major features are:

User Interface: At the center of the Mediator architecture is a user interface based on an interactive visual language tool supporting active concept maps (Gaines and Shaw, 1995a). This user interface, XConMap, was original software developed for the purpose of this thesis. XConMap is a recursive, network aware, object oriented conceptual map editor and browser which is the interface to Mediator1, and the only stand-alone piece of applications software created for this implementation.

Applications: Interaction with Mediator1 occurs predominantly through pre-existing applications software. In the design and creation of Mediator1, the use of as much pre-existing software as possible was a design goal, due to the intended use of Mediator1 by industrial as well as academic partners within the IMS consortium. To satisfy the design objective of utilizing as much pre-existing software as possible, software packages were chosen that operated on multiple platforms, had data sets that were transferable across platforms and could be manipulated by the web.

In Mediator1, this architectural aspect is supported by the following list of applications:

- **XConMap**--XConMap has the dual role of acting as the user interface as well as acting as a supporting member application. When a user accesses an XConMap file, (or a concept map file from another platform) the XConMap editing environment is automatically loaded and the data viewed within it. This is the reason that the application XConMap is referred to as recursive, because it can load, through a node, a concept map and all of its constituent Internet links.
- **Netscape**--Netscape, beside helping to fulfill some of the communication protocols of the lower levels of the Mediator architecture, also acts as a supporting application allowing the display of multimedia files.
- **AutoCAD**--AutoCAD files are stored throughout the Internet and linked to by the nodes in

XConMap. When the links are followed the AutoCAD data is served by the process and communication layer of Mediator, and the AutoCAD application is launched to allow the user to interface with the accessed data.
- **AutoSolid**--In much the same manner as AutoCAD, links to AutoSolid files across the Internet are maintained by the XConMap interface, and when these links are followed, Mediator serves the data to the interface on the desktop by launching an AutoSolid modeling instantiation for the user to interact with the accessed data.
- **FrameMaker**--FrameMaker data is accessed in the same manner as AutoCAD.
- **CNC Tools**--The data from existing CNC editors and applications can also be served up with an HTML interface via Netscape for display and interaction.
- **Process Planning Software**--As with the data from existing CNC editors and applications, the data from existing Process Planning software can also be served through an HTML interface via Netscape for display and interaction.
- **Simulation Software**--The data from existing simulation applications can also be accessed through the Netscape application interface.

Processes and Communications: The process and communications layers of Mediator1 is implemented through the web browsing application, Netscape, which communicates with XConMap through their respective application programming interfaces. The use of the Internet protocols supported by Netscape together with its capability to activate client helper applications greatly simplified the overall implementation of Mediator.

3 Managing the Manufacturing Life Cycle with Mediator

The use of Mediator to manage a distributed concurrent manufacturing activity will be illustrated using example data sets for an engine mounting block. The stages in the product development cycle are: the requirements specification stage, a STEP specification, design drawings (created in AutoCAD), and the manufacturing specifications, which include the Inspection Process Planning and manufacturing Process Planning. These stages are represented in the concept map shown in Figure 2 which is implemented as an interactive diagram in XConMap.

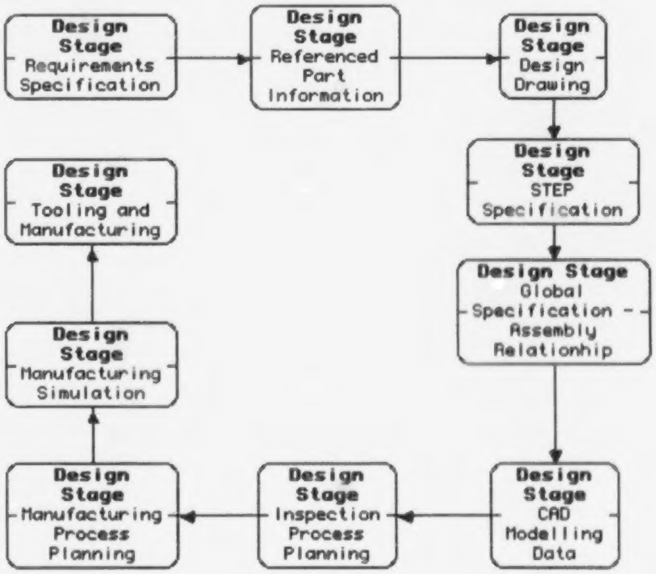

Design Stage Requirements Specification
Design Stage Referenced Part Information
Design Stage Design Drawing
Design Stage STEP Specification
Design Stage Global Specification - Assembly Relationhip
Design Stage CAD Modelling Data
Design Stage Inspection Process Planning
Design Stage Manufacturing Process Planning
Design Stage Manufacturing Simulation
Design Stage Tooling and Manufacturing

Figure 2 Manufacturing life cycle represented in XConMap

In the life cycle of a product, the first stage is the recognition of a need for that product. With the proliferation of planning and organization software on the many platforms available to design and manufacturing engineers, the various means that such a need could be expressed are almost endless. In the case of the sample part the requirements specification is expressed in a document processing application (FrameMaker) that itself supports group access and interaction.

Figure 3 shows the requirements specification in FrameMaker. It was fetched by clicking on the top left node of the concept map of Figure 2 (shown open in XConMap at the top right of Figure 3). XConMap retrieves the URL from its database and passes it to Netscape with a request to fetch the file. When Netscape does so it determines that the file should be opened in FrameMaker and does so. Once this document becomes static rather than dynamic and editable, it could logically be represented in HTML, with its own hypertext links to supporting documentation, references or previous drafts.

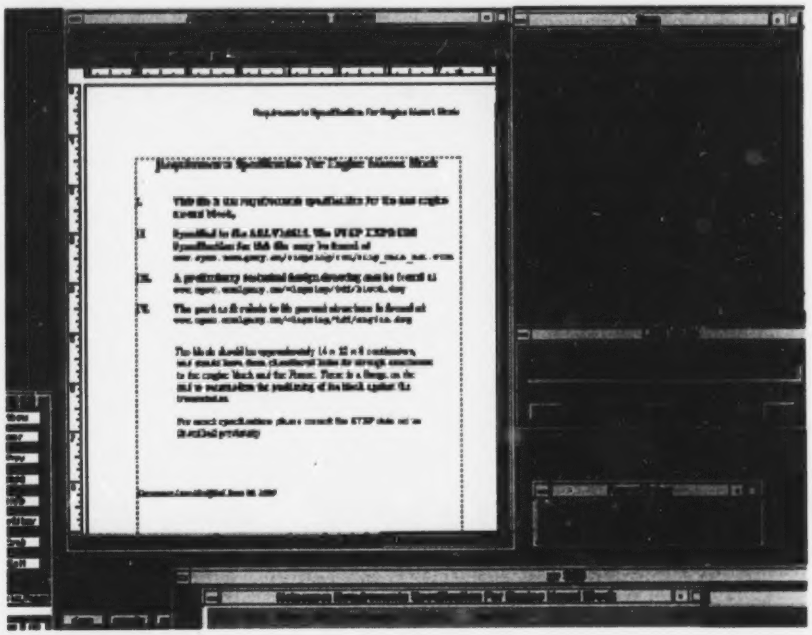


Figure 3 FrameMaker requirements document fetched and opened by XConMap

After the requirements have been proposed and specified through requirements documentation and part referencing, a simple sketch of the product, fitting these requirements, is created. This diagramming can be done in a technical drawing package and in this sample data case, AutoCAD is used. Figure 4 shows the AutoCAD drawing to which the "Design Drawing" node is linked.

BEST COPY AVAILABLE

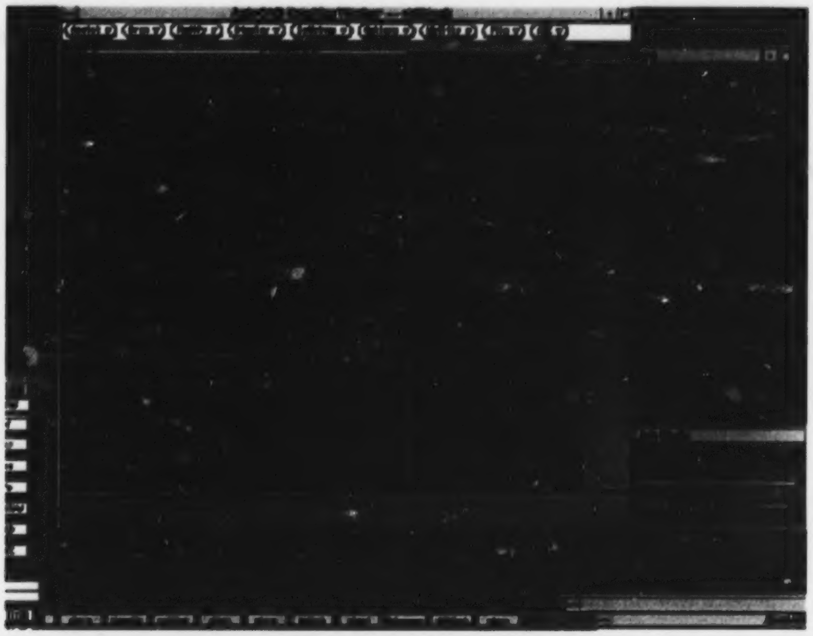


Figure 4 AutoCAD product design drawing fetched and opened by XConMap

STEP (Gu, Zhang and Norrie, 1993) provides an excellent parts description language for product and part information, due to its existence as a global standard. The next stage in the product life cycle is the STEP specification shown in Figure 5. These data are wrapped in HTML tags for easy viewing on the web.

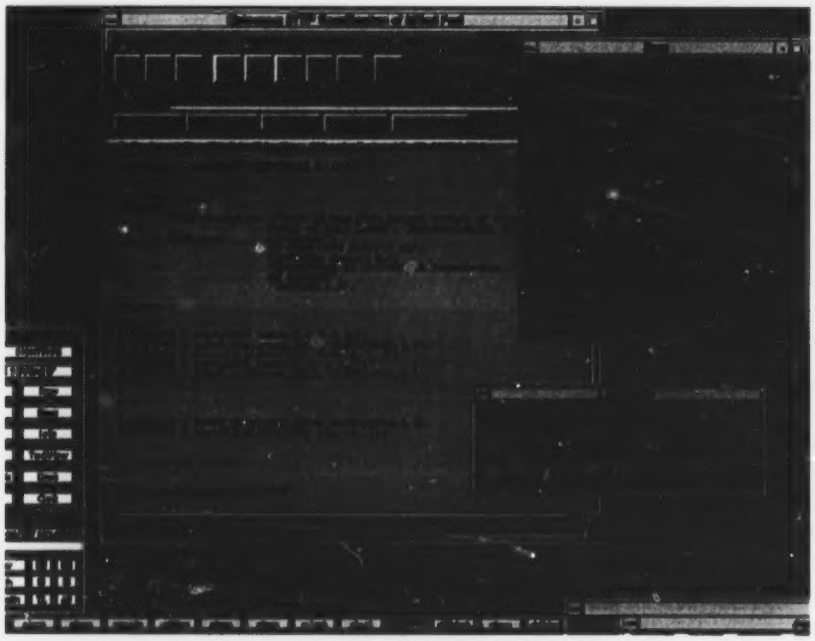


Figure 5 HTML STEP specification fetched and opened by XConMap

XConMap data files can be fetched and opened in XConMap thus allowing concept maps to be linked to

BEST COPY AVAILABLE

one another. Figure 6 shows a process planning concept map fetched across the web by clicking on a node in the life cycle map and then opened in XConMap with active nodes that can themselves be clicked to open other linked files.

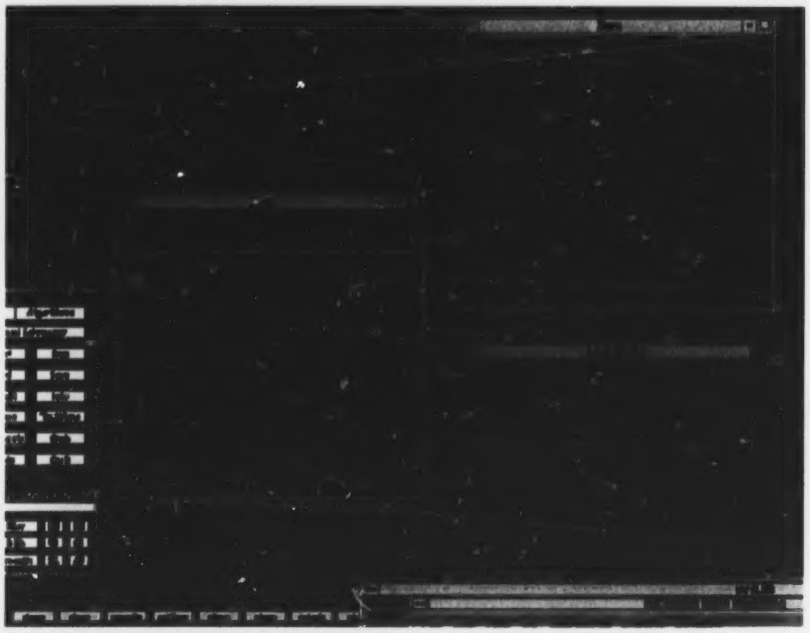


Figure 6 XConMap process planning map fetched and opened by XConMap

5 Conclusions

This article has presented a usable and portable implementation of Mediator1, a Mediator system for the representation of knowledge and the computer support of diverse manufacturing endeavors. The World Wide Web has been used as a knowledge transport mechanism, and the Netscape browser application has been used as a file and data serving agent on the Web. A new application has been created and implemented to provide a visual language interface to Mediator1, called XConMap. XConMap is a recursive, network aware object oriented concept mapping environment built with an eye towards future enhancement, development, and long term cross platform compatibility. This article has shown Mediator1 in use with data sets coordinating inspection process planning, manufacturing process planning, and a complete product manufacturing life cycle.

The existing World Wide Web technologies already provide a major research for the supported of distributed concurrent manufacturing. Further developments are required to fully implement the Mediator design. On major issue is that of awareness in a distributed community--when someone in Canada comes to work in the morning how to they know what their colleagues in Japan and Finland have done overnight without exhaustive search and without requiring those colleagues to take special action to inform. CHRONO (Chen and Gaines, 1996) is a chronological awareness tool that is being developed as a component of Mediator to allow members of a community to maintain awareness of relevant activities.

XConMap and our other concept mapping tools on other platforms have compatible file structures but operate as client helpers. WebMap (Gaines and Shaw, 1995b) is a tool for delivering the concept maps as clickable GIFs to users that do not have a concept mapping client helper. A universal concept mapping tool is under development in Java that will enable concept maps to be used and edited by those with any browser that supports Java. This will provide a visual language interface to Mediator that can be used in a wide variety of applications.

BEST COPY AVAILABLE

The original implementation of Mediator supported synchronous editing at multiple sites. Group brainstorming and discussion is desirable at various stages of the manufacturing life cycle and it will be important to extend web technology to support synchronous interaction including voice communication. A version of Mediator is being developed based on an integration of GroupKit (Roseman and Greenberg, 1992) with World Wide Web that will support synchronous collaboration.

Acknowledgments

This work was funded by Industry, Science and Technology, Canada, as part of the international IMS research program, and by the Natural Sciences and Engineering Research Council of Canada. We are grateful to many colleagues in IMS for their contributions to the Mediator project, in particular, Aldo Dagnino, Peihua Gu, Martti Mäntylä, Tetsuo Tomiyama and Moriki Toyama.

References

Chen, L.L.-J. and Gaines, B.R. (1996). Methodological issues in studying and supporting awareness on the World Wide Web. Maurer, H., Ed. **Proceedings of WebNet96**. This conference. Charlottesville, VA, Association for the Advancement of Computing in Education.

Gaines, B.R. and Norrie, D.H. (1994). Mediator: information and knowledge management for the virtual factory. **SIGMAN AAAI-94 Workshop: Reasoning about the Shop Floor**. pp.30-39. Menlo Park, California, AAAI.

Gaines, B.R. and Norrie, D.H. (1995). Knowledge systematization in the international IMS research program. **Proceedings of 1995 IEEE International Conference on Systems, Man and Cybernetics**. pp.958-963. New York, IEEE.

Gaines, B.R., Norrie, D.H. and Lapsley, A.Z. (1995). Mediator: an Intelligent Information System Supporting the Virtual Manufacturing Enterprise. **Proceedings of 1995 IEEE International Conference on Systems, Man and Cybernetics**. pp.964-969. New York, IEEE.

Gaines, B.R. and Shaw, M.L.G. (1995a). Concept maps as hypermedia components. **International Journal Human-Computer Studies 43**(3) 323-361.

Gaines, B.R. and Shaw, M.L.G. (1995b). WebMap: concept mapping on the web. **World Wide Web Journal 1**(1) 171-183.

Gu, P., Zhang, Y. and Norrie, D.H. (1993). Object-oriented product modelling for computer-integrated manufacturing. **Integrated Computer-Aided Engineering 1**(2) 17-28.

Kusiak, A. (1993). **Concurrent Engineering: Automation, Tools, and Techniques**. New York, Wiley.

Norrie, D.H. and Gaines, B.R. (1995). The learning web: A system view and an agent-oriented model. **International Journal of Educational Telecommunications 1**(1) 23-41.

Roseman, M. and Greenberg, S. (1992). GroupKit: a groupware toolkit for building real-time conferencing systems. **Proceedings of the Fourth Conference on Computer-Supported Cooperative Work**. pp.43-50. New York, Association for Computing Machinery.

Tomiyama, T. (1992). The technical concept of IMS. RACE Discussion Paper, No. RA-DP2, Research into Artifacts, Center for Engineering, The University of Tokyo.

KSI Page

gaines@cpsc.ucalgary.ca 16-Jul-96

On-the-Fly Web Pages Minus the HTML Files

Gerry Lebing, U.S. Geological Survey, USA, glebing@usgs.gov

Maura Hogan, U.S. Geological Survey, USA, mhogan@usgs.gov

Wesley Gurney, U.S. Geological Survey, USA, wgurney@usgs.gov

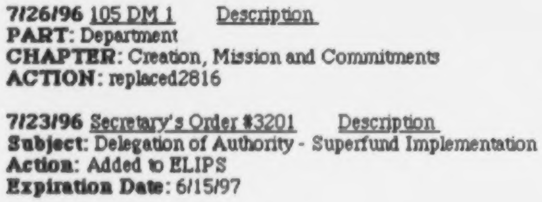
The demands of a data-driven distributed hypermedia environment drastically alter the development and production of Web products. When these demands are combined with emerging publication priorities of authoring once and distributing via numerous media for numerous audiences, it becomes apparent that coding documents with HyperText Mark-up Language (HTML) is limited and singular in purpose.

A new paradigm for Web development is emerging in response to a constantly evolving information environment. Data-driven, responsive, dynamic Web products will be landmarks in this digital terrain. Web spaces that provide visitors the ability to customize the presentation and scope of information will further enhance this environment. These challenges can be addressed integrating databases into on-the-fly production of Web pages.

Creating on-the-fly Web pages requires information providers to question the process and taxonomy of information presentation, storage, and use. The role of HTML moves from a form of communication (i.e., how the content is packaged) to a vehicle of communication (i.e., how the content is exchanged and delivered). The role of the Web page as an autonomous lexia becomes paramount in order to achieve the multiplicity of context and purpose that can exist in a hypertext "docuverse." A dynamic, data-driven Web site can be summarized as a "Web site that can construct Web pages on the fly, based on customer inputs, data from databases, and collections of components that are all stitched together at run time to build custom pages to answer users' questions and serve their needs and interests" [Scannell 1996]. Although these demands seems immense and unwieldy, emerging technologies will facilitate the automation and maintenance of these dynamic Web environments. For example, the number of applications currently supporting Web/database access is already substantial--48 at a recent count--and Web/database discussion groups already exist [1].

In the on-the-fly model, a database field can be viewed as a lexia as its content becomes a building block instead of part of a predetermined Web page. These fields can be queried, retrieved from the database, and then sent to the Web browser based on a template. The template, created with a common gateway interface (CGI) program, outlines how to present the results in HTML. The same information can be used concurrently as building blocks for multiple pages, thus eliminating the need to replicate the same content in more than one document.

This on-the-fly approach has implications for production and usability. In terms of production, the task of maintaining and updating information stored in a database as opposed to HTML files is more manageable and efficient. Each update, change, or correction is made once in the database, and all of the pages using that information are automatically and instantaneously as current as the database. This capability eliminates the need of locating multiple HTML documents where the change needs to be made and manually updating each file. This approach also supports the goals of computer-support cooperative work [2]. Content specialists working in a distributed hypermedia environment can update their portion of a database as needed, and these changes are immediately available to users of that information. An example of this type of distributed work is ELIPS, the Electronic Library of Interior Policies for the Department of the Interior (DOI), a collection of standards for the publishing of guidelines and regulations [3]. The content of the entire ELIPS manual exists in multiple databases that can be queried, searched, and generated on the fly. Content specialists use internal password protected fill-in forms to make changes to the databases. For example, WordPerfect files uploaded to the database are automatically linked and indexed. Because ELIPS is data driven, as a change is made in the database, a series of automated tasks take place. When a change is made, the "Chronological List of Changes to ELIPS" is automatically updated to reflect the change.

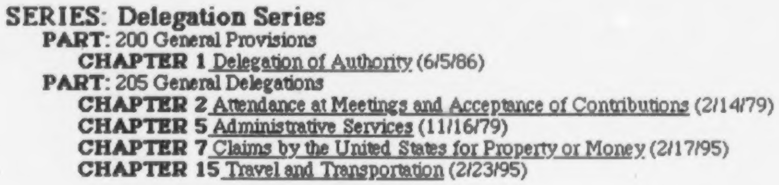


7/26/96 105 DM 1 Description
PART: Department
CHAPTER: Creation, Mission and Commitments
ACTION: replaced2816

7/23/96 Secretary's Order #3201 Description
Subject: Delegation of Authority - Superfund Implementation
Action: Added to ELIPS
Expiration Date: 6/15/97

Figure 1. Example of changing entries in the "Chronological List of Changes to ELIPS."

At the same time, the "Table of Contents" is updated to reflect the last revision date of a particular series, part, or chapter.



SERIES: **Delegation Series**
- **PART**: 200 General Provisions
  - **CHAPTER 1** Delegation of Authority (6/5/86)
- **PART**: 205 General Delegations
  - **CHAPTER 2** Attendance at Meetings and Acceptance of Contributions (2/14/79)
  - **CHAPTER 5** Administrative Services (11/16/79)
  - **CHAPTER 7** Claims by the United States for Property or Money (2/17/95)
  - **CHAPTER 15** Travel and Transportation (2/23/95)

Figure 2. Example of changing entries in the "Table of Contents."

Users have an immediate sense that the information is a dynamic document and that the information they are accessing is current. Using this on-the-fly approach facilitates the maintenance of Web products, exemplifies distributed searching across multiple databases, reduces the need to edit multiple documents containing the same content, partially automates the delivery of the final Web product, and ensures that the content is immediately as up to date as the database.

In terms of usability, the information in database fields can be reused for multiple applications. Each new use requires only the creation of a new template, not the duplication of the content. For example, the USGS URL Registry (limited to USGS access) is a database that collects metadata for suites of USGS Web pages. URL, associated URLs, Title, Contact, Theme, Geographic Extent, Discipline, Audience, Keywords, and Expiration Date are the fields that are requested through a fill-in form. Multiple products can be created from this database, including internal pages catologing content by State and public pages with a thematic organizing approach. Additionally, the database can be queried to locate specific information--for example, hazard information (theme) in Virginia (geographic extent) in terms of hydrology (discipline)--and to generate reports that are geared to help Webmasters maintain pages (i.e., locate pages that have expired).

The business advantages of creating dynamic, data-driven Web pages include reducing the duplication of content across multiple pages, increasing the efficiency of maintaining rapidly changing content, and increasing the potential number of employees who can contribute to the maintenance of a database. The on-the-fly Web model requires information providers to consider information in terms of an information taxonomy that supports multiple contexts. Equally important is the concept that the same information can have multiple uses for multiple audiences. Using different on-the-fly templates to reach these diverse audiences answers the challenge of the new paradigm: Web pages that are data driven, responsive, and dynamic.

- [1] For information on Web/database discussion groups, see http://cscsun1.larc.nasa.gov/~beowulf/bbs/.
- [2] For information on CSCW, see http://arirang.snu.ac.kr/seminar/ddblab/sykim/cscwdir.html.
- [3] For information on ELIPS, see http://elips.doi.gov/.

References

- Scannell, Ted (1996). Steve Jobs Interview. InternetUser: http://www.softbank.co.jp/softbank/publishing/internetuser/jobs/english.html

Merz: Creating Personal and Shared Spaces on the World Wide Web

Sören Lenman
Invited researcher
CID, NADA, KTH, S-100 44 Stockholm, Sweden.
lenman@nada.kth.se

Henry See
Invited researcher
CITI. 1575 Boul. Chomedey, Laval, Québec, Canada H7V2X2.
hsee@citi.doc.ca

Michael Century
Director of research
CITI, 1575 Boul. Chomedey, Laval, Québec, Canada H7V2X2
mcentury@citi.doc.ca

Bruce Pennycook
Professor
McGill University, Faculty of Music, Montréal, Québec, Canada
brp@sound.music.mcGill.ca

Abstract: Merz is an environment which supports knowledge work, personal information management and collaboration on the WWW. As a user gathers information by browsing and querying, a local database of metadata is built, serving as a basis for creating Merzbaus: visual representations of documents and links. The user interface is characterised by filtering and semantic zooming. Instead of having to classify objects into hierarchies and folders, users can successively add keywords to the metadata records for objects and links and create multiple, filtered views of information. Categorised objects can be moved to parallel, more organised Merzbaus, e.g., corresponding to specific projects. Users can personalise Merzbaus by rearranging icons, creating links, and by entering hand-written marks. Collaboration is supported by means of publishing Merzbaus to a WWW-MOO server for other users to access. Merz is implemented in JAVA.

Introduction

Merz is a new kind of user environment for the WWW, which supports personal information management, knowledge work, and collaboration. Merz is being developed on Sun workstations, using the JAVA multiple-platform language [Sun 1996]. Collaboration in Merz is implemented using MOO-technology [Curtis 1991].

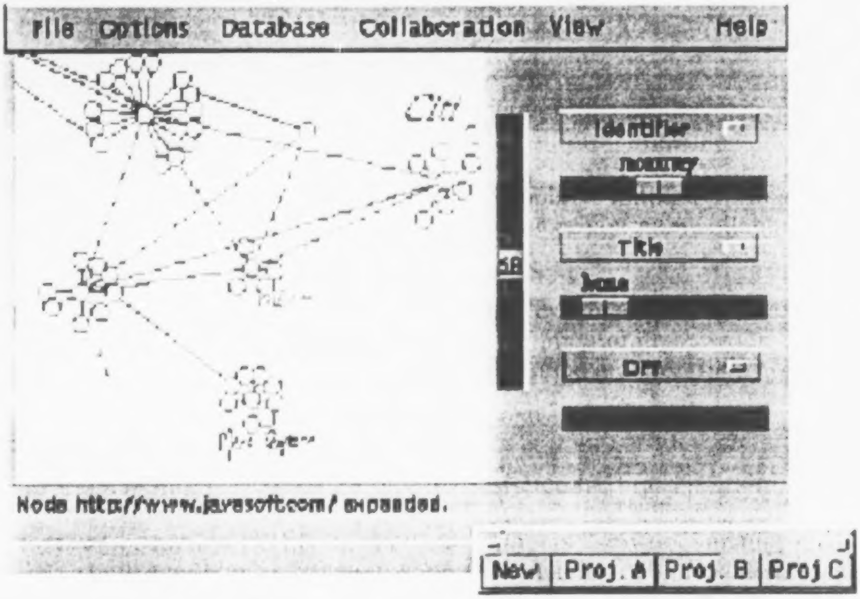
As a tool for personal information management and knowledge work, Merz primarily addresses usability problems like information overload, disorientation, and difficulties in organising information, which are common in hypermedia systems [Nielsen 1995]. In Merz, the key approach to helping users overcome these problems is by visualisation of information. This is not an uncommon approach. However, Merz is unique in three respects, first, that users can personalise visualisations, e.g., by laying out node and link elements by hand and by adding hand-written annotations, drawings and marks, second, that users have access to multiple views of information, e.g., by rapidly and seamlessly changing metaphors and semantic filters, third, that gradual organisation of information is supported, as opposed to having to immediately organise it into hierarchical structures. The emphasis on such personal, creative and dynamic aspects of visualisation motivated the use of the concepts "Merz" and "Merzbau", which were borrowed from the German artist Kurt Schwitters [Schmalenbach 1967]. A Merzbau is a construction put together of bits and pieces found in the world (see [Merz) below).

As a tool for collaboration Merz is aimed at situations where groups need to share visualisations, access other groupmembers' annotations, follow their trails through information, and collaboratively develop information structures. Merzbaus can be shared with other users for them to explore, modify and return to the originator. Thus, the collaborative aspects of Merz support building information structures at a conceptual, "project-organising" level, rather than, e.g., the more common collaborative creation of hypermedia documents.

Below we give a simple example of Merz, followed by an overview of previous, related work, a description of the current status of the implementation, and some future development and research issues. An earlier version of Merz has been described in [Lenman and See 1995]

Merz example

A simple Merzbau and a preliminary user interface is illustrated in [Fig. 1]. A similar, JAVA demo is available on the WWW [Merz 1996]. The circles represent HTML-documents and the light grey lines represent hyper-links embedded as HREF tags in the documents. New documents are added, e.g., by browsing the WWW. Embedded links are automatically extracted and represented, e.g., as the circular patterns in [Fig. 1].



Node http://www.javasoft.com/ expanded.

Fig. 1 A Merzbau and a preliminary user interface

Zoom and pan are primary interface operations. The vertical slider to the left of the window indicates the current zoom-level, and can be used to control the zoom directly. However, the main method for zooming and panning is by means of the mouse in combination with different keys. Double-clicking on an object gives access to various actions: viewing the object itself with a WWW-browser, annotating, editing, removing the visual representation of the object from the Merzbau, etc. Different kinds of objects support different actions.

In [Fig. 1] the user has entered some personal notes and marks by drawing directly in the space. Also, two links (dark grey) have been drawn between objects by the user, to indicate a relation. The nature of the relation can be specified in annotations or keywords, which are stored in the metadata record for the link.

The horizontal sliders to the right in [Fig. 1] control filters, based on semantic characteristics of objects. What sliders represent can be selected from a menu, accessible from the slider. In [Fig. 1], a single object which has "mcentury" in the identifier and "home" in the title matches the filter and is highlighted. Filters tuned to, e.g., different projects a user is working on, can be saved and restored.

The metaphor by which a Merzbau is viewed can be changed, e.g., to an ordered list of file names instead of a spatial view, depending on the task at hand. In [Fig. 1] the metaphor is very simple: objects on a 2D surface. As understanding of information pieces grows they can be represented in parallel, more organised Merzbaus, e.g. corresponding to specific projects, concepts, or ideas.

Lastly, Merzbaus can be shared with other users by exporting (or importing) them by means of the "collaboration" menu. Shared Merzbaus can be modified and re-exported, e.g., as part of work within a project-group.

Background and related work

Merz

Kurt Schwitters (1894-1947), who invented Merz in 1919, was a writer, publisher, artist, designer, architect, ad-man, and performer. Merz served Schwitters as a quasi-brand name linking together his paintings, poems, plays, sculptures, and architectural constructions: "The word Merz denotes essentially the combination, for artistic purposes, of all conceivable materials, and technically, the principle of the equal evaluation of the individual materials. A perambulator wheel, wire-netting, string and cotton wool, are factors having equal rights with paint." [Schwitters 1919]

The WWW promises, but does not today provide this kind of suppleness, this capability to work freely with heterogeneous materials, if not for artistic purposes, for purposes of gaining and developing knowledge. As a set of tools designed to help individuals better organise, integrate, and understand the heterogeneous materials of cyberspace, the Merz application is not so much an aid to efficiency, but to coherence. It is a tool to deal with conditions of unbounded proliferation of elements -uncatalogable in principle- which exists already on the WWW.

The discovery of new relationships between things that Merz aims to support, and that is implicit in Schwitters' Merz principle, bears a close resemblance to the concept of "abduction", by the philosopher C. S. Peirce [Buchler 1955]. Neither a deduction from formal premises, nor an induction of general rules from instances, abduction can be understood as the logic of creativity. To our knowledge there is only one other Web based research project, the Mental World Browser, which starts from similar concerns [Kubota 1996]. Another similar project, although not concerned with the Web, aims at developing computer tools that support engineers in articulation of unstructured, nebulous, fragmented ideas that are typical during certain phases of creative activities [Hori 1993].

Knowledge work

Three particular characteristics describe knowledge work: annotation and note-taking, spatial layout and manipulation of information pieces, and the difficulty to enter information into classification schemes before it has been understood. In note-taking, the personalisation that results from personal marks and annotations seem to be important for the creative process of understanding. Thus, computer tools should be aimed at capturing and reproducing the appearance of marks and notes rather than at interpreting them [Kidd 1994].

Other researchers have also observed that knowledge workers use physical space as temporary holding patterns for information pieces they cannot yet categorise or decide how they might use. They often prefer to deal with information by creating physical piles of paper, rather than immediately categorising it into specific folders [Mander, Salomon and Wong 1992]. Even when we find it difficult to classify items, or to express how or why they are interconnected, we are often able to meaningfully arrange them in space [Marshall and Shipman 1995].

Thus, computer tools should not force users to classify information in order to store it for later use. Making sense out of information is a process; annotation, manipulation of information items, and experimenting with spatial arrangements are important activities in this process. Although it might seem self-evident that information cannot be classified until it is understood, most current systems require exactly this: users are provided only with an hierarchical file system, which requires immediate labelling if information is to be stored meaningfully. In contrast, Merz is specifically aimed at supporting personalisation, spatial manipulation and deferred, gradual classification of information.

Visualisation

Using spatial models and metaphors for accessing and managing large repositories of information is not a new idea. The history of visualisation is a long one [Tufte 1990] [Rieber 1995], and in human-computer interfaces spatial metaphors have been used extensively. There are numerous examples of visualisation of large scale file systems and libraries, e.g., the recent work at Xerox PARC [Robertson, Card and Mackinlay 1993]. Visualisation of the WWW is also common, e.g., the Navigational View Builder [Mukherjea and Foley 1995], Hyper-G [Andrews, Kappe and Maurer 1995], and HyperSpace [Wood, Drew, Beale and Hendley 1995]. There are also recent developments towards 3D visualisation on the WWW, as reflected in the VRML standard and applications [Pesce 1995]. However, the approach to visualisation in Merz is different, because it emphasises not visualisation per se, but that different users require different views of information, depending on preferences, circumstances and tasks, and that changing the view should be user-controlled, rapid and seamless.

In Merz the main components of a view are semantic zoom, filter, and metaphor. Semantic zoom refers to how much detail about the content is revealed. For example, depending on the scale at which the content is viewed, an item or a link can be represented by an icon, a title, a summary, or the full page, e.g., as rendered by a standard Web browser. The main purpose is to close the gap between navigational views and close views, thus avoiding the phenomenon of being "lost in hyperspace". [Waterworth and Chignell 1991]. E.g., while reading a document, a user only needs to quickly zoom out for reorientation, and then zoom back in, or pan and zoom to a different document. An earlier, advanced development using this principle is Pad++ [Bederson, Stead and Hollan 1994].

The filter component of a view refers to what content is shown. For example, a user could choose to see only documents related to "scientific visualisation". All other information in the view could be filtered out completely, or shown in a dimmed rendition. Such a mechanism supports working with information in flat information structures, as an alternative to, or in parallel with a hierarchical organisation. The background to this approach is Shneiderman's Dynamic Queries [Shneiderman 1994] and the recent development IVEE [Ahlberg and Wistrand 1995].

Metaphor refers to the frame of reference which is used to present information to the user. A variety of metaphors can be applied to organise the same selection of information. For example, a collection of information objects can be viewed as an alphabetical list, as folders and subfolders in a desktop metaphor, or as places, villages, and houses in a 3-D virtual world. The purpose of using metaphors is to reduce complexity in organising information and to aid the user's memory by exploiting prior knowledge from different domains [Carrol, Mack and Kellogg 1988]. However, forcing a single, detailed metaphor can be both counterproductive and cumbersome, and restrict inventiveness and creativity [Waterworth and Chignell 1989]. In Merz, users can choose to view information organised by different metaphors, depending on the task at hand.

The Merz application

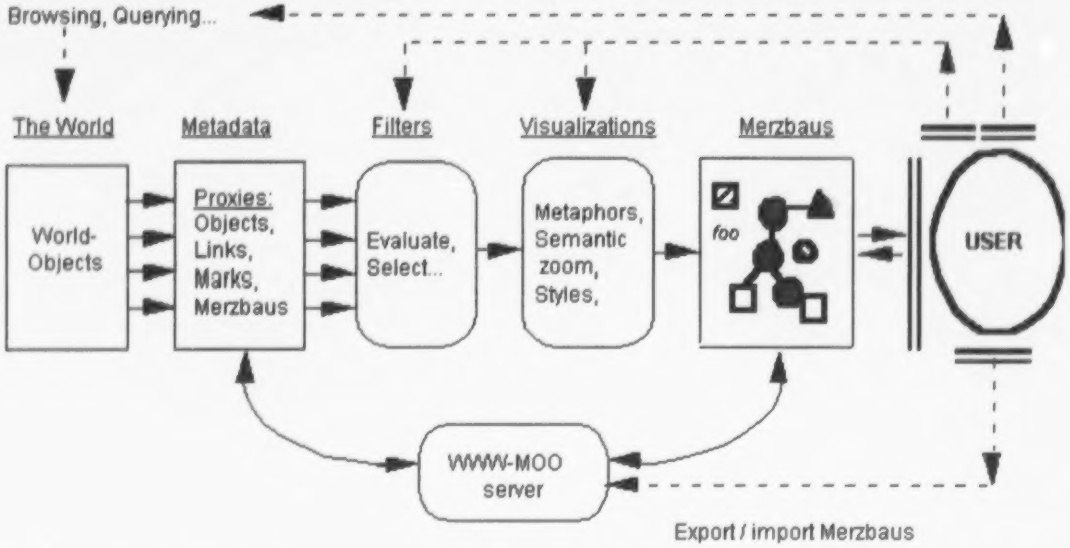


Fig. 2 Overview of the main components in the Merz implementation. Parallel, double-width lines indicate user interfaces.

The structure of Merz, which is implemented as a JAVA application, is shown in [Fig. 2]. Information is brought into Merz by browsing or querying the WWW, or by automatic processes, "agents". However, the latter aspect has not yet been implemented. Accessing a new information item (a world-object) creates a database record, a proxy, which holds metadata about the object in a local database, maintained for each user. This database is central to the Merz environment, both for supporting visualisation and for the collaborative aspects. The base fields of proxies conform to the Dublin Core Set [Godby and Miller 1996], such as title, kind of object, address, keywords, geographic location, etc. There are also additional fields, e.g., to hold keywords and annotations that a user might add. There are four kinds of proxies: proxies for world-objects, proxies for links, proxies for user-created marks, and proxies for Merzbaus.

When an HTML-document that contains embedded links is accessed, it is automatically parsed, and a separate proxy for each link is created. Thus, in Merz links are bi-directional and treated as first-class objects, which can be given keywords and annotated, as can any other object.

A Merzbau proxy holds references to user-created collections of world-objects, which themselves have proxies in the database. Merzbau proxies also have one or many associated view specifications, which in the current implementation are stored as separate files, and which hold information about how the Merzbau should be visualised.

Proxies (world-objects) are represented to the user through a filtering process, which selects information from the database according to criteria specified by the user. Information from the filter is passed to a visualisation process, which assigns representations to the selected information within the framework of a metaphor, according to some rules of assignment. The resulting view is in most cases visual, but currently we are also working on auditory representations of information. Parallel representations are possible, e.g., different windows can show different views onto the same Merzbau.

Collaboration

In the current phase of the project it was most straightforward to implement collaboration through a central server. This server had to be internet compliant, allow user registration and password protection, permit user group assignments, support document protection and archiving, and it had to be fully programmable and sufficiently flexible to permit modifications and additions. We choose to use a variant of Lambda MOO [Curtis 1991]. Members of the Merz software team had previously developed WWW - MOO software, e.g., [Virtual Film Festival 1995], based on the hypertext extensions to Lambda

MOO [Meyer, Blair & Hader 1995].

MOOís are object-oriented databases which include a rich collection of primitives, object-classes, and functions ("verbs") applicable to multi-user, networked collaboration. The integration of MOO technology with the WWW through HTML declarations permits complex hyper-media sites to be constructed. For the Merz project, however, a different scheme was required. First, the PERL routines which provide Web - MOO correspondence had to be replaced with JAVA applications. Second, the MOO would not provide interactive on-line services for users but, rather, would remain on-line transparently once the user had made a valid connection. After such a connection, a user may "put" a Merzbau, identified by a unique name, into the MOO. Any other authorized user attached to the server can then retrieve the Merzbau by name, and start making changes to it. Thus, the MOO functions as a background database server for the storage, archiving and retrieval of Merzbauís.

Each Merzbau is identified by a unique name and is owned by a unique user. Named Merzbauís and users are, in fact, uniquely numbered objects in the MOO database and can be accessed by any of the MOO procedures. Ownership privileges are assigned to objects according to the privileges of the user. That is, if user N has group access to all the Merzbaus submitted by users A, B, and C, then N will be able to ìgetî the Merzbau, i.e., the Merzbau proxy and all its associated proxies and view specification(s). If user N re-submits this Merzbau it will be stored by a new name and assigned new privileges. A very simple scheme exists to format the client-side Merzbau data into an ASCII stream of name-value pairs. Companion client-side and server routines have been implemented in JAVA which manage the connections and the put and get routines.

Thus, at the current stage, several users can put, get, modify, and re-put Merzbaus as new objects in a secure, fully protected database. The sequencing of requests and read/write accesses are handled by the MOO core software. While Lambda MOO databases were not initially designed for tasks of this nature, all of the criteria listed earlier were realized using this novel combination of JAVA and lambda MOO.

Future developments

There still remains some work concerning the basic JAVA application, especially with regard to aspects of visualisation which involve metaphors. Another concern is the user interface: the existing interface is only a point of departure for the interface development, which will take place through rapid prototyping and formal usability testing.

Adding autonomous processes to aid search and monitoring of content on the WWW is a next step. Some preliminary work in this respect has been carried out concerning the use of neural networks, and how the user should communicate with such processes in a visualisation environment. Another, ongoing line of development concerns the previously mentioned "sonification" of information: using sound to convey meaning [Pennycook Breder and Dawkins 1996].

The system for collaboration implemented to date will be extended to also permit direct person-to-person communication. This will include many of the standard features of the Web - MOO, as well as real-time audio conferencing using low-bandwidth speech codec systems such as TrueSpeech [DSP Group 1996]. A further initiative will be to replace the central MOO server with a distributed, client-based system. This development is necessary to fully realise the kind of seamless collaboration we have in mind for users of Merz.

References

[Ahlberg and Wistrand 1995] Ahlberg, C. & Wistrand, E. (1995). IVEE: An Environment for Automatic Creation of Dynamic Queries Applications. CHI'95 Conference Companion, Denver, Colorado, 15-16.

[Andrews, Kappe and Maurer 1995] Andrews, K., Kappe, F. & Maurer, H. (1995). Hyper-G and Harmony: Towards the Next Generation of Networked Information Technology. CHI'95 Conference Companion, Boston, MA, 33-34.

Bederson, Stead and Hollan 1994] Bederson, B.B., Stead, L. & Hollan, J.D. (1994). Pad++: Advances in Multiscale Interfaces. CHI'94 Conference Proceedings Boston, MA, 315-316.

[Buchler 1955] Buchler, B. (1955). Philosopical Writings of Peirce. New York: Dover Publications.

[Carrol, Mack and Kellogg 1988] Carrol, J.M., Mack, R.L. & Kellogg, W.A. (1988). Interface Metaphors and User Interface Design. In Helander, M. (Ed.) Handbook of Human-Computer Interaction. North Holland: Elsevier Science Publishers B.V., 67-85.

[Curtis 1991] Curtis, P., (1991). The Lambda MOO Programmer's Manual, Xerox PARC. Available via FTP from parcftp.xerox.com/pub/MOO/.

DSP Group 1996] TrueSpeech (1996). DSP Group, Inc. 3120 Scott Boulevard Santa Clara, CA 95054-3317. http://www.dspg.com/

[Godby and Miller 1996] Godby, J. & Miller, E (1996). A Metalanguage for Describing Internet Resources Using the Dublin Core Element Set. Proceedings of INet'96, Montréal, QC, Canada. Available from http://www.isoc.org/.

[Hori 1993] Hori, K. (1993). Computer-supported Articulation of the Tacit Dimension. Toward Dynamics of Knowledge in Engineering. Proceedings of the 1st Int. Symposium of Research Into Artifacts, Tokyo, Japan, 100 - 107.

[Kidd 1994] Kidd, A. (1994). The Marks are on the Knowledge Worker. CHI 94 Conference Proceedings, Boston, MA. 186-191.

[Kubota 1996] Kubota, A. (1996). The Abduction Machine Project. RACE, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, 153, Japan. http://artifacts.race.u-tokyo.ac.jp/hn/AbM.html

[Lenman and See 1995] Lenman, S. & See, H. (1995). Personal Information Spaces. Poster Proceedings of the 4th International WWW-conference, Boston, MA. 56-57.

[Mander, Salomon and Wong 1992] Mander, R., Salomon, G. & Wong, Y. (1992). A Pile Metaphor for Supporting Casual Organisation of Information. CHI'92 Conference Proceedings, Amsterdam, the Netherlands, 627-634.

[Marshall and Shipman 1995] Marshall, C.C. & Shipman, F.M. (1995). Spatial Hypertext: Designing for Change. Communications of the ACM, 38 (8), 186 - 191.

[Merz 1996] The Merz Project (1996). CITI, 1575 boul. Chomedey, Laval (Québec) H7V2X2. http://merzbau.citi.doc.ca/merz.html

Meyer, Blair & Hader 1995] Meyer T, Blair, D. & Hader, S. (1995). WAXweb. A MOO-Based Collaborative Hypermedia System for WWW. Computer Networks and ISDN Systems, 28 (1&2), 77-84

[Mukherjea and Foley 1995] Mukherjea, S & Foley, J., D. (1995). Visualizing the World-Wide Web with the Navigational View Builder. Computer Networks and ISDN Systems, 27 (6), 1075-1087

[Nielsen 1995] Nielsen, J. (1995). Hypermedia and Hypertext. The Internet and Beyond. Boston, MA: AP Professional.

[Pennycook Breder and Dawkins1996] Pennycook, B., Breder E. & Dawkins, K. (1996). The Merz Soundscape. McGill University, Faculty of Music, Montréal, Québec, Canada. http://www.music.mcgill.ca/~brp/merzscape.html.

[Pesce 1995] Pesce, M. (1995). VRML: Browsing and Building Cyberspace. New Riders / MacMillan Publishing

[Rieber 1995] Rieber, L.P. (1995). A Historical Review of Visualization in Human Cognition. ETR&D, 43 (1), 45-56.

[Robertson, Card and Mackinlay 1993] Robertson, G.G., Card, S.K. & Mackinlay, J.D. (1993). Information Visualization Using 3D Interactive Animation. Communications of the ACM, 36 (4), 57-71.

[Schwitters 1919] Schwitters, K. (1919) Die Merzmalerie, Der Sturm, (10 (4), quotation from Dietrich, D. (1993) The Collages of Kurt Schwitters, Cambridge University Press.

[Schmalenbach 1967] Schmalenbach, W. (1967). Kurt Schwitters. New York: Harry N. Abrams Inc.

[Sun 1996] Sun Microsystems Inc. (1996). Java - Programming for the Internet. Sun Microsystems, Inc., 2550 Garcia Ave., Mtn. View, CA 94043-1100 USA. Http://java.sun.com

[Shneiderman 1994] Shneiderman, B. (1994). Dynamic Queries for Visual Information Seeking. IEEE software, 11 (6), 70 - 77.

[Tufte 1990] Tufte, E.R. (1990). Envisioning Information. Connecticut: Graphics Press.

[Virtual Film Festival 1995] The Virtual Film Festival (1995). 5505 Saint-Laurent, Suite 4104, Montreal, Québec Canada H2T1S6. http://www.virtualfilm.com.

[Waterworth and Chignell 1989] Waterworth, J.A. & Chignell, M.H. (1989). A Manifesto for Hypermedia Usability Research. Hypermedia, 1 (3), 205-234.

[Waterworth and Chignell 1991] Waterworth, J.A. & Chignell, M.H. (1991). A Model for Information Exploration. Hypermedia, 3 (1), 35 - 57

[Wood, Drew, Beale and Hendley 1995] Wood, A., Drew, N., Beale, R. & Hendeley, B. (1995). HyperSpace: Web Browsing with Visualization. Third International WWW Conference Poster Proceedings, Darmstadt, Germany, 22 - 26.

Acknowledgements

Sören Lenman was working as an Invited researcher under the National Exchange Program at CITI, 1575 Boul. Chomedey, Laval, Québec, Canada H7V2X2 during this project.

Aspects of Large World Wide Web Systems

J. Lennon
Department of Computer Science,
Private Bag 92019, Auckland, New Zealand - H. Maurer
Graz University of Technology,
Schießstattgasse 4a, A-8010 Graz, Austria

Abstract:

In this paper we consider some of the challenges associated with many World Wide Web systems, particularly in their handling of large quantities of multimedia data, and ask whether we can look forward to better systems in the not too distant future. In particular, we look at various aspects of hypermedia data models including new methods for link management.

Introduction

The World Wide Web (WWW) must, on many counts, be rated a success: the number of servers, the number of users, the amount of hype published daily all around the world. However, in this paper we look at some of the consequences of using systems that have large numbers of unstructured nodes with considerable numbers of links [Conklin, 1987], particularly on distributed systems such as WWW [Kappe, 1995]. Although hardware related problems such as inadequate bandwidth obviously still exist, we concentrate on more fundamental data model issues, especially those associated with link structures.

Internet Problems

It has, however, become quite apparent that a number of fundamental changes in Internet systems are necessary if they are to continue to grow and to be used for large professional projects. We will begin by outlining just a few of the more common problems experienced by users:

- Dangling links. As we shall elaborate in Section 2, most of the commonly-used systems cannot guarantee any degree of referential link integrity. All too often users are faced with some form of ``file doesn't exist" message. It is particularly frustrating to discover that documents that were successfully accessed one day have disappeared the next!
- Loss of orientation. This well known problem of large systems is one consequence of the fact that links in many WWW systems are unidirectional, have no type, and thus cannot support users by providing location cues such as local area maps or tours.
- Inappropriate user interfaces. Although browsers such as Netscape are suitable for many users, and a variety of tasks, there are however notable exceptions. For example, specific interfaces need to be carefully designed for systems such as public terminal viewers where novice users need guidance so that they can find the information they want as efficiently as possible. On the other hand, experienced users also need better support. Currently, developing HTML material is often a two screen process one screen running the editing application while the other screen runs a specific browser.
- Difficulty in finding specific material and knowing whether the found information is the ``best" available, and whether there is more of interest.
- Difficulty of gaining an overview of the material.

In the face of problems such as these, many new features for the next generation of hypermedia systems are needed, six of which are discussed here in detail.

Link Management

Although in many respects link management is the crux of hypermedia system design it is, unfortunately, the weakest aspect of many modern systems. Indiscriminate use of links, particularly in large distributed systems, has led to tangled and mismanaged webs, frustrating for system managers and users alike. In Section 3 we discuss the structuring of data in ways that can minimise the use of links with no loss of functionality. In this section we discuss why it is important that links should be:

- Not embedded in the documents themselves (see Section 2.1)
- Bi-directional (see Section 2.2)
- Typed (see Section 2.3)

Why Not Embedded?

The fact that links are imbedded in the documents themselves is a major limitation of first generation systems. We believe that the only practical way to keep referential integrity, particularly in large hypermedia systems, is to store and maintain the links in a separate link database. At the very least, links to documents that are removed should be automatically un-highlighted - as is currently implemented in the Hyper-G system [Maurer et al., 1993a, Kappe et al., 1993, Kappe et al., 1994]. Other interesting options are outlined in [Lennon and Maurer, 1996]. The Hyper-G system also supports ``open" anchors which remain un-highlighted and inactive until such time as the link is automatically resolved when the ``missing" documents are added to the system.

Another significant consequence of having non-embedded links is that it enables anchors to be placed in non-standard objects such as movies, postscript files, and even the texture maps on 3-D objects [Andrews et al., 1995a, Maurer, 1996].

Finally, as we shall elaborate in Section 5, links that are stored in a separate database may have user rights associated with them. This fact alone has far-reaching consequences.

Why Bi-directional?

The second requirement for links is that they be bi-directional. Bi-directional links have the following advantages:

- Minimisation of the dangling link syndrome. When links are bi-directional, all the documents that point to any particular document can be located. In large systems this provides the only practical means for notifying authors about impending dangling links.
- Automatic link maintenance. As we have mentioned, the Hyper-G system automatically de-activates links that would point to a deleted document. This is only possible in a system with bi-directional links stored in a separate link database.
- Navigational maps. Much better navigational maps can be provided using a structured system with bi-directional links. Several graphical browsers such as Hyper-G are providing history trails and three-dimensional maps of the hyperspace [Andrews et al., 1994].
- Statistics. The only way of collecting statistics such as ``who points to this document?" is by following links back to their sources.

Why Typed?

Future systems should support a set of predefined link types, as well as user-defined link types. The document entitled ``What is a Link Type & How Many Are Enough" contains an interesting summary of a USENET discussion [Trickel1996, 1996]. The discussion originated from a posting that commented on a quote from Conklin, ``Trigg describes over 80 such links" [Conklin, 1987]. The lively discussion supported the view that users should be able to define their own link types. NoteCards [Halasz, 1988], for example, ``supports an infinite number of link types" [Trickel1996, 1996]. It is important that there be equally good support for link types in WWW systems [Lennon and Maurer, 1996] - particularly for link filtering, by the type of document or by link author.

Structuring data

The unstructured and tangled nature of links in many WWW systems is a source of frustration to users and administrators alike. As van Dam said, people ``got linkitis" [van Dam, 1988]. There is certainly no significant data model for WWW, and in many ways first generation systems may rightly be likened to novice programs that used too many GO TO statements. A case for developing hypermedia systems without links is made in [Maurer et al., 1994].

Second generation systems such as Hyper-G [Andrews et al., 1995a] allow system administrators to implement systems which reflect the data's inherent structure. Hyper-G [Maurer, 1995b] organises data objects into clusters [Kappe et al., 1993], that are grouped into collections, which in turn may be grouped together in a pseudo-hierarchical manner - as described in the next section. Hyper-G thus imposes structure on top of a flat file database. Most significantly, it provides users with an alternative to link browsing. Users navigate up and down hierarchies and thus can follow the inherent structures.

In the article ``Hyper-G Organises the Web" the author states, ``It [Hyper-G] can organise the mass of unstructured data and unmanageable hyperlinks" [Flohr, 1995].

The HM-Data Model

The HM-Data Model, developed at Graz Institute of Technology, addresses each of the following problems that are usually associated with first generation systems:

1. Loss of orientation encountered by users.
2. Tedious editing of links where, as we have mentioned in Section 2, deleting documents can result in dangling links.
3. Loss of visible semantic structure when sets of nodes are combined with other sets [Andrews et al., 1995b].

The HM-Data model supports the structured browsing of hypermedia databases [Maurer et al., 1993a].

The databases consist of a set of structured collections called `S_Collections` (or just collections for short) [Maurer et al., 1995]. An `S_Collection` encapsulates both a chunk of multimedia data (content), and a particular internal structure. The internal structure consists of a ``pseudo-hierarchical" group of other collections. We use the term ``pseudo-hierarchical" because in fact the structure is a directed acyclic graph of parent and children nodes, where a child may belong to more than one parent. This means that any document can belong to multiple collections - without duplication of content.

The HM Data Model provides a number of predefined subclasses of `S_Collection` (see Figure 1) as follows:

1. An Envelope where all members are inter-linked
2. A Folder which is essentially an ordered list

3. A Menu which is a simple hierarchical structure
4. Freelinks where members may be arbitrarily connected

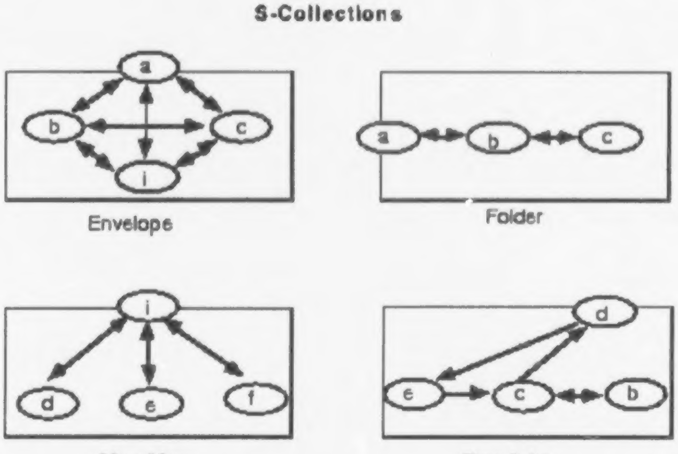


Figure 1: HM Data Model Collections (Redrawn from [MSS93])

An imposed condition in the HM-Data Model is that all links must be encapsulated within `S-Collections`. Browsing is achieved by the means of the three operations ACCESS, ZOOM-IN, and ZOOM-OUT. When any `S_Collection` is ACCESSed its content is displayed. ZOOM-IN opens a collection to display the underlying content. ZOOM-OUT takes the user back to the collection that was accessed just prior to the most recent ZOOM-IN. We thus have a structure, which does not have to rely on explicit links, where users can browse up and down the hierarchies while keeping a ``well defined sense of the current location" [Maurer et al., 1993a]. Also, since at any moment of time a stack of open `S-collections` may exist it is easy to generate a meaningful graphical trace.

It is important to note that although any collection can be a member of other collections this does not mean that the data has to be duplicated: pointers may be used. Also recursive membership is both possible and meaningful.

Besides supporting multi-metaphor browsing [Maurer et al., 1993a] the HM-Data model incorporates a powerful query mechanism in the navigational paradigm [Maurer et al., 1993b] thus providing users with an integrated environment.

Meta-data

One of the little addressed deficiencies of most WWW hypermedia systems is the lack of meta-data. Meta-data is information about a document such as its title, author, date of insertion into the system, date of last modification, access rights, expiry date, etc. It can also include user-defined information such as keywords and summary data. This information can be used for a wide variety of purposes, just a few of which we mention here:

- Meta-data information can form a basis for better searches).
- Necessary statistics can be collected and categorised by date, user, etc.
- As we discuss in Section 5 chunks of multimedia data can be linked to form personalised documents. Using access rights, the customised documents can be made available to specific groups of people at specific times.
- Perhaps most significantly in these days of fast moving change, only by having meta information available can systems be more readily upgraded to new and/or better systems. It is interesting to note, for example, how readily a system such as Gopher can be ported to Hyper-G [Maurer,

1995b].

Customisation of Documents

Undoubtedly, one of the thorniest set of problems that must be solved by the next generation of hypermedia systems is that of copyright issues. This is a particular problem of concern for authors of courseware. In previous papers [Maurer, 1995a, Lennon and Maurer, 1996] we have elaborated the idea of creating documents, not by copying from other books and journals, but by linking to publications for which we have library-like user rights. We suggest that ``information brokers" will supply information rights, at a cost, to various organisations [Maurer, 1995a]. Within ``information hubs", say on a LAN with a hypermedia server, authors can create links to ``chunks" of information as suggested by Figure 2. The customised document can then be made available to selected groups of users by setting appropriate link and document access rights.

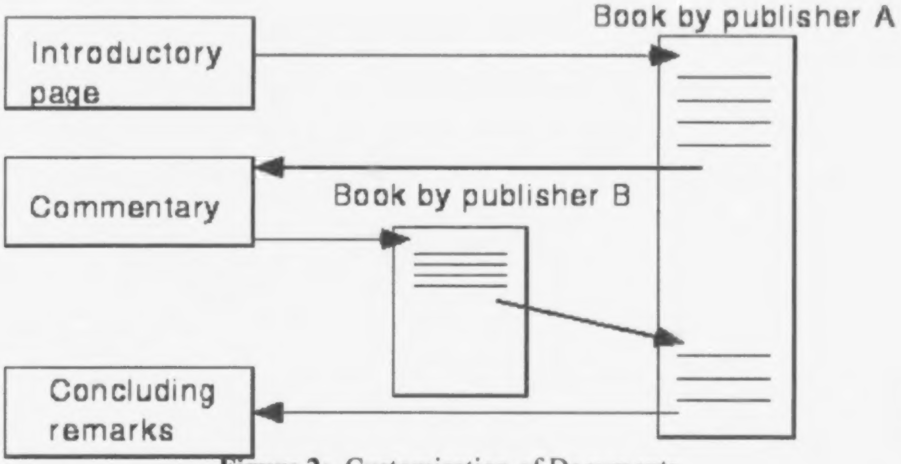


Figure 2: Customisation of Documents

Almost certainly users will access the information on a first come first served basis as in traditional libraries. It will be the information manager's responsibility to see that users are not locked out by insufficient copies - the greater the demand the more copies the provider must supply and pay for.

An important consideration is how we should control the process of users adding source anchors to library documents. It may be necessary, for example, that authors give permission for information providers to set up the link-access rights so that groups of authorised users can add links to write-protected documents. We realise, however, that to protect systems from being overloaded considerably more research needs to be done to determine just how the user rights need to be managed.

Sequencing

As a final example of the significant advantages of having a separate link engine, consider the process of maintaining sequential lists of URLs in most WWW systems. For example, consider the steps that must be taken if a new Document named Maurer has to be inserted into a list already containing Lanier and Nelson (see Figure 3).

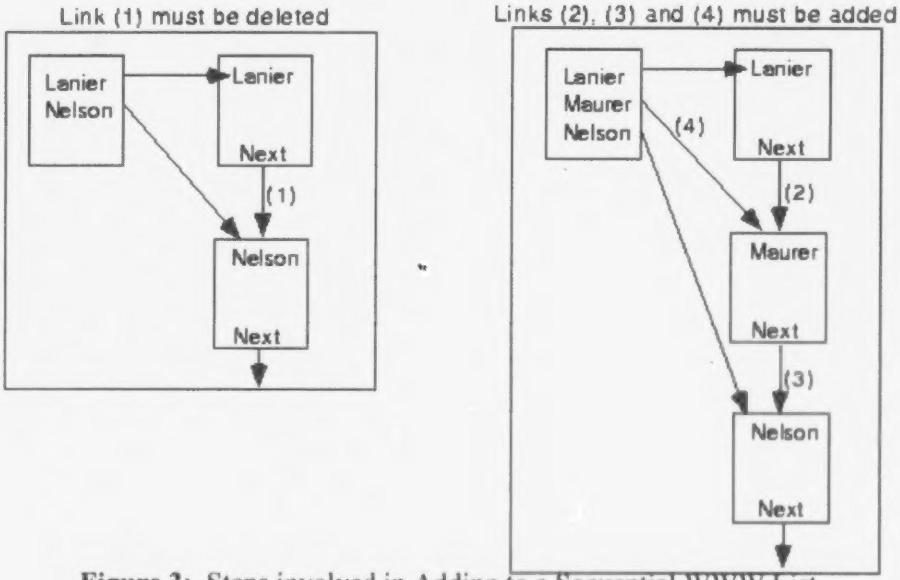


Figure 3: Steps involved in Adding to a Sequential WWW List

As shown, the document Maurer has to be inserted first. Then the document representing the alphabetic list Lanier, Nelson,... has to be updated to include Maurer; finally, the link labeled (1) between the Lanier and Nelson documents must be deleted, and the new links (2),(3), and (4) created. In sharp contrast, in a Hyper-G system only a single operation is needed, i.e., the document Maurer is simply inserted into an appropriate collection, and the system does all the rest when the ``sequence" attribute is set.

Conclusion

In this paper we considered some of the WWW problems that arise from having large amounts of unstructured data in flat file databases. We then considered the advantages of systems which have a separate link engine that supports bi-directional and non-embedded links. We looked forward to hypermedia systems which integrate e-mail, discussion groups, CSCW and video conferencing, supported by easily defined link types. Finally we looked at ways of structuring documents to minimise link maintenance while supporting features such as customisation and sequencing of documents.

References

Andrews et al., 1994
Andrews, K., Kappe, F., and Maurer, H. (1994). The Hyper-G network information system. In *Special issue: proceedings of the Workshop on Distributed Multimedia Systems, Graz, Austria.*

Andrews et al., 1995a
Andrews, K., Kappe, F., Maurer, H., and Schmaranz, K. (1995a). On second generation hypermedia systems. In *Proc. ED-MEDIA 95 - World Conference on Educational Multimedia and Hypermedia*, pages 75-80. AACE.

Andrews et al., 1995b
Andrews, K., Maurer, H., and Scherbakov, N. (1995b). Semantic modeling of object-oriented hypermedia databases. In Blumenthal, B., Gornostaev, J., and Unger, C., editors, *Human-Computer Interaction, EWHCI'95*, pages 121-134. Springer, LNCS 1015.

Conklin, 1987

Conklin, J. (1987). Hypertext: An introduction and survey. *IEEE Computer*, 20(9):17-41.

Flohr, 1995
Flohr, U. (1995). Hyper-g organizes the web. *Byte*, pages 59-64.

Halasz, 1988
Halasz, F. (1988). Reflections on notecards: Seven issues for the next generation of hypermedia systems. *Communications of the ACM*, 31(7):836-852.

Kappe, 1995
Kappe, F. (1995). A scalable architecture for maintaining referential integrity in distributed information systems. *Journal of Universal Computer Science*, 1(2).

Kappe et al., 1994
Kappe, F., Andrews, K., Faschingbauer, J., Gaisbauer, M., Maurer, H., Pichler, M., and Schipflinger, J. (1994). Hyper-G: A new tool for distributed hypermedia. In *Proc. Distributed Multimedia Systems and Applications Conference*, pages 209-214. ISMM-ACTA Press.

Kappe et al., 1993
Kappe, F., Maurer, H., and Sherbakov, N. (1993). Hyper-G - a universal hypermedia system. *Journal of Educational Multimedia and Hypermedia*, 2(1):39-66.

Lennon and Maurer, 1996
Lennon, J. and Maurer, H. (1996). Flexible link architectures in hypermedia systems. In *Accepted for Proc. ICCE 96*. AACE.

Maurer, 1995a
Maurer, H. (1995a). Hypermedia systems as internet tools. In van Leeuwen, J., editor, *Computer Science Today*, pages 608-624. Springer, LNCS 1000.

Maurer, 1995b
Maurer, H. (1995b). On some new aspects of networked multimedia systems. In Bartosek, M., Staudek, J., and Wiedermann, J., editors, *SOFSEM'95: Theory and Practice of Informatics*, pages 315-333. Springer, LNCS 1012.

Maurer, 1996
Maurer, H., editor (1996). *Hyper-G The Second Generation Web Solution*. Addison-Wesley.

Maurer et al., 1993a
Maurer, H., Kappe, F., Scherbakov, N., and Srinivasan, P. (1993a). Structured browsing of hypermedia databases. In *Proc. HCI'93*, pages 51-62, Vienna. Springer.

Maurer et al., 1994
Maurer, H., Philpott, A., and Scherbakov, N. (1994). Hypermedia systems without links. *Journal of Microcomputer Applications*, 17:321-332.

Maurer et al., 1995
Maurer, H., Scherbakov, N., and Schneider, A. (1995). HM-Card: A new hypermedia authoring system. *Journal for Multimedia Tools and Applications*, 1(3).

Maurer et al., 1993b
Maurer, H., Scherbakov, N., and Srinivasan, P. (1993b). A new hypermedia data model. In *Proc. DEXA'93*, pages 685-696, Prague, Czech Republic. Springer, LNCS 720.

Trickel1996, 1996
Trickel1996 (1996). http://www.win.net/ tpi/hypertext/usentlnk.htm. HTTP.

van Dam, 1988
van Dam, A. (1988). Hypertext'87 keynote address. *Communications of the ACM*, 31(7):887-895.

Jennifer Lennon jlennon@hmu.auckland.ac.nz

SADS : Generic Web Access to Space Data

Thierry Levoir,
CNES, CT/TI/PS/TD
18, avenue Edouard Belin
31055 Toulouse Cedex, FRANCE
E-mail : levoir@cst.cnes.fr

Rémi Bastide
L.I.S., Université Toulouse I,
Place Anatole-France,
31042 Toulouse Cedex, FRANCE
E-mail: bastide@cict.fr

Abstract: This paper is an experience report from the SADS project which was conducted at the French Space Agency (CNES) in Toulouse, France. SADS is the acronym for "Système d'Accès aux Données Spatiales", Space Data Access System. The project was aimed at offering Web access to the huge volume of scientific data coming from earth observation satellites that the CNES maintains. The project was also meant to serve as a testbed for Web development environments and techniques that were new at that time at CNES, and to study how those techniques could be integrated in the Software Engineering practices and rules currently in use. The SADS project presents several interesting aspects and problems that were to be solved : the enormous volume of data that are to be accessed and eventually transferred to the user's site, the special structure of this data, which may vary according to the various experiments, and the fact that further evolution of the system had to be planned in this first prototype.

Introduction

This paper introduces the SADS project, which aims at offering to the scientific community a Web access to the data gathered in space experiments, stored and maintained by the CNES. This projects uses some common Web development techniques (e.g. CGI gateways to relational databases), and other not-so-common ones (e.g. access to large data storage and retrieval systems). The paper also focuses on organizational and software engineering issues : We report our experience in deploying Web development techniques, which are still new and rapidly evolving, in conventional software developments procedures, and also detail the solutions we have chosen to ensure long term maintenance and evolution of the project.

Goals of the SADS

Any large scale scientific project (such as solar or earth observation satellites) usually develops a data access system, to allow research teams to access to data stemming from their own instrument. These systems are conceived on a per-project basis, and are designed for users who know perfectly the structure of the data in question.

The continued used of these specific systems after the end of the mission is hindered by two major difficulties :

- after the exclusive utilization phase of data by the Principal Investigator (PI), the data has to be made available to a larger scientific community that does not know the detail of the embarked instruments. However, the specific data access systems were not conceived in the first place for those non-initiate users.

- the necessity, today recognized, to preserve space data in the long term, can in no case have for consequence the duty to maintain in functioning state all systems developed by projects, and this for several decades.

One of the CNES' goals is to make sure that the data gathered during spatial missions can be preserved for undetermined amounts of time, and that this data remains accessible to a large community of users from the scientific community. These constraints have led the CNES to conceive and to put in operation a system allowing these users to identify available data, to select data elements potentially interesting for their research work, and to access and retrieve this data.

The emerging popularity of the World-Wide-Web and its acceptance as a de-facto standard in the scientific community have conduced the CNES to base its data access solution on the HTTP protocol and related technologies such as HTML and CGI gateways, instead of providing a proprietary solution.

Web Access to Huge Amounts of Data

One of the special characteristics of the SADS system is that is aims at offering access to an amazingly huge amount of scientific data maintained in long term physical storage and conservation service set up by CNES and known as STAF (*Service de Transfert et d'Archivage des Fichiers* - file transfer and storage system), which operates in a heterogeneous environment. This system was conceived and put in operation before the start of the SADS project, and had to be integrated in our design solution.

Architecture of the SADS System

The SADS is based on the following components [Fig. 1]:

- A metadata management system based on a relational model. This metadata is the data dictionary of the scientific data that is made available through the SADS. It consists in the description of the data from the different scientific missions stored in the SADS (e.g. record types and formats) and of some related data that can be used by the scientists to make queries (e.g. the occurrence of unique events such as solar eruptions or satellite malfunctions). The metadata is stored in an Oracle database and queried through CGI gateways.
- A physical data storage function compatible with the IEEE mass storage reference model. This is the purpose of the STAF.
- A metadata access function allowing users to identify available data, to display items available for browsing and to fill out data queries. The querying may be aided by the visualization of scientific images, which helps the user in spotting particularly interesting time intervals. A sophisticated user interface allows users to incrementally construct their queries through thumbnail images browsing and other visual querying techniques. This is handled by a software package providing also for system protection, data protection and implementation of generic data extraction mechanisms.
- Data order processing and user delivery functions, which allow registered users to store typical queries within the SADS, or to automatically transfer query results to their own site.

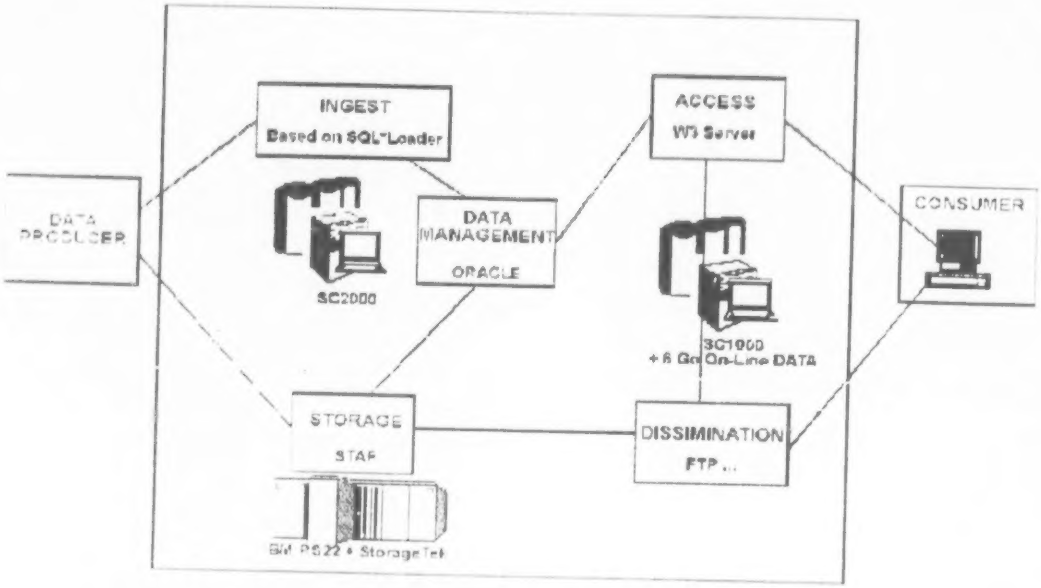


Figure 1: Architecture of the SADS system

Technical Solutions

Due to the potentially huge volume of data that may be retrieved after a user request, it is usually impossible to provide the answers to the query in real time. We thus had to devise other solutions, combining the use of electronic mail and FTP software, allowing for a more "asynchronous" transfer mode. According to the user's profile and to the volume of data retrieved by his query, the data may be stored in a temporary area of the SADS, waiting for the user to download it by FTP, or automatically uploaded to the user's site if it is equipped with a FTP server. It is even conceivable to revert to surface mail for sending the data (which was the usual transfer mode before SADS), since the retrieved volume of data routinely reaches tens or hundreds of megabytes. In any case, the user is informed of the availability of his query results by Email.

Introducing Web Development in an Existing Organization

SADS was conducted in the CNES as a prototype project, aimed both at asserting the feasibility of a Web solution for such a long-term and technically challenging project, and to explore the Software Engineering problems related to the use of such a new and rapidly evolving technology.

Most notably, we had to define how the new programming techniques and languages related to the HTTP protocol (HTML, scripting languages, public domain tools) could be validated and integrated in the very strict CNES quality insurance plan. This led us to define development rules and testing techniques for Web applications, so that the SADS, consisting in static HTML pages, dynamically created pages, relational database modeling and programming and had-hoc gateway software development could be validated with the same strictness as any conventional software development.

We were also committed to strictly adhere to existing standards, such as standardization of dates and times in compliance with CCSDS recommendations [CCSDS 1990]. We were able to develop a general program for the extraction of data corresponding to one or more time frames defined by the user from one or more files. This program makes it possible to extract only that data which corresponds strictly to the time frame requested by the user [Fig. 2]. The extraction function is entirely independent of the archive's structure:

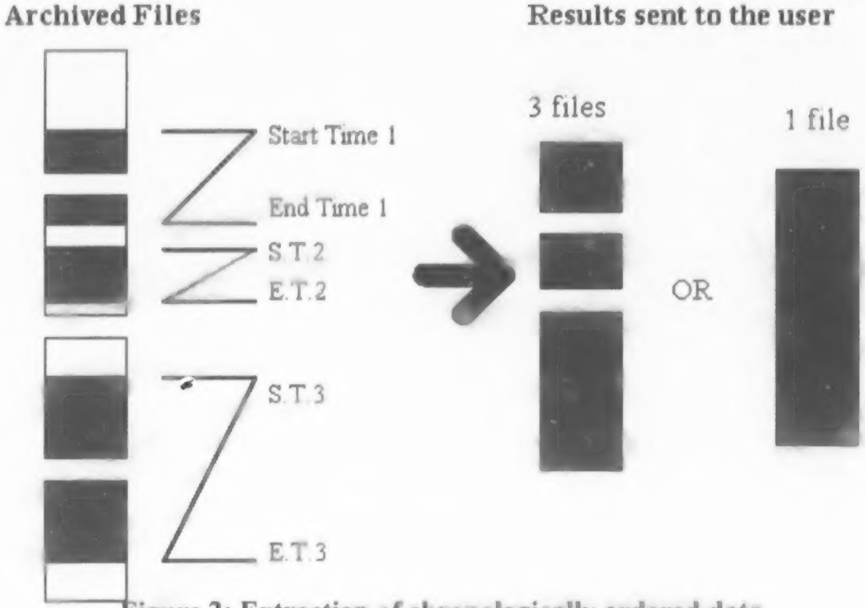

Archived Files
Results sent to the user
Start Time 1
End Time 1
S.T.2
E.T.2
S.T.3
E.T.3
3 files
1 file
OR

Figure 2: Extraction of chronologically ordered data

We also use standardised data descriptions based on the EAST language (Enhanced Ada SubseT) [CCSDS 1995]. This is a formal language around which certain general tools have been or are currently being developed. The use of this tool involves two stages:

- a first stage in which the user selects the fields in which he is interested. A hierarchical tree representation of the different data fields is constructed on the basis of the EAST description. Using this representation, users identify and mark the fields in which they are interested through a HTML form-based interface.
- a second stage in which data is extracted from an archive and then 'filtered' so as to preserve only those fields requested by the user [Fig. 3].

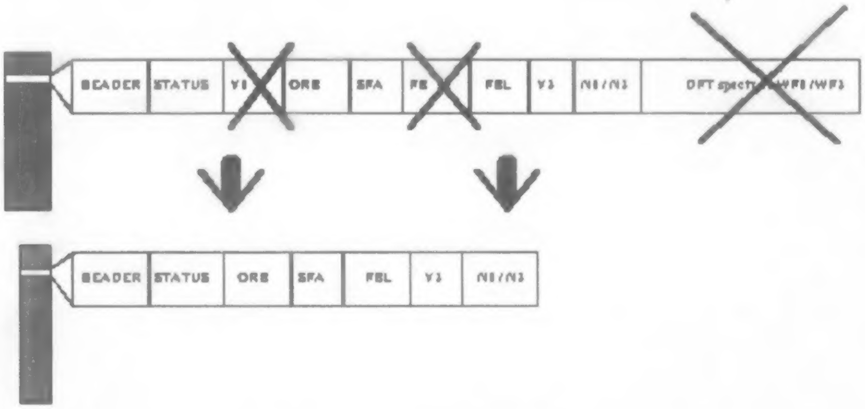

HEADER
STATUS
V1
ORB
SFA
FE
FEL
V3
N1/N3
DFT spectr
WF1/WF3
HEADER
STATUS
ORB
SFA
FEL
V3
N1/N3

Figure 3: Field extraction

Special care was taken in the design of the SADS to ensure that the proposed solution would allow for an easy evolution, since eventually past and future space mission will store their data in this host structure. It was thus one of our design goals to propose a solution flexible and generic enough to accommodate for yet unforeseen storage requirements.

- The metadata database model was designed to accommodate for any data record structure, whether fixed or variable length. It is also able to deal with various date representations.
- The metadata stored in the database needs to be maintained by the SADS administrator, but, for each mission stored in the SADS, a set of information pages has to be provided by the mission's Principal Investigator. These pages are provided in standard HTML format, but, to ensure uniformity of presentation among the various missions, a template structure for those pages has been designed. These templates ensure homogeneous presentation and navigation structure between missions, while preserving the freedom for the PI to design the pages' contents at will.

Conclusion and Future Work

The SADS prototype is currently undergoing intensive tests within the scientific community, mainly in the field of Plasma Physics. At the present time, the system offers access to 32 different sets of data acquired during 5 space missions (INTERBALL, Sweden VIKING, ISEE1, VOYAGER, GEOS). Approximately 1 giga-bytes of space data are accessed by the prototype by month.

A critical analysis of the limitations of this approach is being carried out. The analysis focuses in particular on system adaptability to all types of data, (whereas the data currently stored in the SADS is mainly structured and queried according to time periods). This analysis forms the basis for an object-oriented model of a more general prototype known as AGDS (Accès Généralisé aux Données Spatiales, Generalized Access to Space Data). Through the use of inheritance in the metadata model, we expect to make easier the insertion in the system of data of more various nature.

In the same time the CDPP project (Plasma Physics Data Center), is under way, to offer an operational access system to Plasma Physics Data.

References

[CCSDS 1990] CCSDS (1990). Time code Format, CCSDS 301.0-B-2, blue book, issue 2, April 1990.

[CCSDS 1995] CCSDS (1995). The data description language EAST specification, CCSDS 644.0-R-1, Red book, November 1995.

[CCSDS 1996] CCSDS (1996). Reference Model for Archival Information Services, June 1996, Version 4.0 (Available from http://www.gsfc.nasa.gov/nost/isoas/overview.html)

[Huc 1995] C. Huc (1995). A strategy for the long term preservation of space mission data, JTS95 Symposium, London, January 27-29, 1995

[Huc et al. 1996] C. Huc, T. Levoir, M. Nonon-Latapie (1996). Long-term archiving and data access: modelling and standardization (published at the Fifth Nasa/Goddard Space Flight Center Conference On Mass Storage Systems And Technologies, September 17-19, 1996)

[Larzul et al. 1996] B. Larzul, D. Minguillon, P. Mazal (1996). EAST Technology for an automated processing of Space Data (to be published at the SPACE OPs symposium, Munich, November 1996)

HELP-EXCHANGE: An Arbitrated System for a Help Network

Paul Losleben
Stanford University
CIS-110, MC-4070
Stanford, CA 94305
415/725-3620
plosleben@stanford.edu

Alice S. Tull
Escondido Village
Apt 36D
Stanford, CA 94305
tull@cs.stanford.edu

Erming Zhou
P.O. Box 3804
Stanford, CA 94305
zhou@cs.stanford.edu

Abstract: The problem of obtaining help over the Internet continues to be a perplexing challenge, especially as the Internet grows. Most existing systems do not scale well as the number of users increase. The Help-Exchange system has been developed in an attempt to facilitate person-to-person connections, especially across disciplines in a way that improves as the number of users increases. The system is based on a taxonomy of disciplines and arbitrates a dialog between users by matching questions to experts. The paper describes the features of the system and the results of early testing.

Background

About five years ago, a set of activities were spawned from the Microelectronics Systems Education Conference [Borriello 91] aimed at improving the effectiveness of microelectronics education. One fledgling idea which happened to not be funded at the time sought to build a network of educators and researchers who could assist each other through an email-based system designed to connect individuals with questions to other individuals who might have an answer. This idea was suggested by an attendee of the conference and sought to solve problems of scaling in existing systems.

The rapid growth of the VLSI community in the early 1980s was largely due to extensive interaction between researchers and educators in Electrical Engineering and Computer Science departments at a relatively small number of universities. The sharing of knowledge, tools and designs was impressive, but the well-known individuals in the community quickly became overburdened with requests for information. In a much broader sense, newsgroups [Kantor & Lapsley 86, Horton & Adams 87] have been enormously popular and successful, but the traffic volume is often too high to be useful, even with partitioning by topic into hundreds of separate newsgroups. There are also many systems now in use to connect individuals with questions or problems to experts who are either paid to provide advice or are volunteers in formal organizations of one form or another. These generally go by the name of "HelpDesk" and are familiar to those of us who have tried to get customer support on a product we have just purchased. None of these approaches scale well.

These "person-to-person" approaches to providing information are quite different from and complement the enormous body of work that has been underway for at least two decades to locate static information over electronic networks. Valuable information often exists only in the minds of experienced individuals and even if the information is available in machine form, human reasoning is often very valuable in identifying and utilizing context in searching for information.

Thus, in recognition that human networks can be extremely valuable, the idea suggested the need for a system that would help a community of individuals to seek help from one another. We call it a "Help-Exchange" system.

The Arbitrated Help-Exchange Concept

The concept is based on three very simple ideas: 1) that everyone is an expert in some domain, 2) that expertise in one domain might be traded for expertise in another and 3) that a mechanism is needed to spread the demand for help so that no single individual is overloaded. In its simplest form, the concept is like match-making coupled with a barter system. The system implementing Help-Exchange matches people with questions with other people who might provide answers, keeps score, and balances the load of questions on any user. It is arbitrated in that it serves as a fair and neutral intermediary between the users, assuring anonymity if desired. There is also an underlying assumption of good will on the part of all users. Beyond that, it is mostly transparent in the resulting dialog.

Description of the Help-Exchange System

A prototype of the Help-Exchange system was built during the 1994/5 academic year as a task in the National Research Enterprise project [Losleben 96] at Stanford University. It was first demonstrated in August 1995, alpha tested with a small test group during the late summer and is presently in extended beta test with a larger community of users at Stanford and MIT. The World Wide Web provided key enabling technology in the form of a standard communication protocol, universally available client software, easy-to-develop user interfaces and direct access to the underlying relational data base. The system uses a parallel email communication path to alert users and incorporates a mailbox analogy inside the system to classify correspondence.

Taxonomy of Disciplines

Like most information systems, some mechanism is needed to classify knowledge and to label the classes with descriptive titles which carry sufficient connotation to communicate the content of the class. We elected to use a hierarchical taxonomy consisting of simple English words or phrases to describe the categories. This is admittedly prone to error in interpretation due to subtle differences in semantic meaning. In addition, the taxonomic organization is developed at the discretion of the system administrator and therefore also biased by that individual's view of how expertise might be organized. We chose to not address these difficult research issues [Gruber 93] in the interest of building a simple tool which is easy to use across a broad taxonomy.

Most interactions with the Help-Exchange system begin with the taxonomy. By convention, the higher levels of the taxonomy represent more general knowledge while the lower areas represent more specialized knowledge. New users register their area(s) of expertise by selectively expanding the taxonomy until they reach a level of specialization that they feel appropriate to their expertise. A user may register in multiple areas if they feel that they have expertise in those areas. Clearly these are self-appointed experts and that is a weakness of the system. Still, this is consistent with the assumption of good will on the part of all users.

When a user seeks to pose a question to the system, the taxonomy is used to identify the most likely category of knowledge. The user is provided with an indication of the number of experts who are available (in another words, who have registered and haven't fulfilled their obligation to the system). Of course, if no experts are available for a category, that category may not be selected and the user may elect to submit the question at a higher and more general level of expertise.

When an expert receives a question which is not classified properly, the expert can refer the question to an expert in another category, again by using the taxonomy. The expert can also refer the question to another expert in the same category if he/she does not feel qualified to answer.

The Barter System

Since the Help-Exchange system is entirely dependent on the good will of the users, some mechanism is needed to assure that load balancing is fair and that there is a clear limit to the workload that any user might sign up for. Provision is made for the user to increase or decrease this limit at any time as long as the balance between the right to ask questions and the obligation to answer questions is kept. The system utilizes a simple barter system: answers are considered as credits for the right to ask questions. In an ideal world, one might wish for an even exchange, the right to ask one question in exchange for providing one answer. Unfortunately economic systems are not so simple and some "float" is required to assure economic stability in the system. Initially, we ask that each user agree to answer two questions for each one that they ask. As we gain experience with the system, this may be changed.

Anonymity

Both the person asking a question and the person supplying an answer may elect to be anonymous. We felt that this was important in a cross-disciplinary dialog. In newsgroups, the "newbee" is often discouraged from asking what might be interpreted as "dumb" questions. Of course, no-one can be an expert in all domains, so there is no good measure of what might constitute an improper question. Unfortunately, since newsgroups are designed to broadcast to a large audience, an innocent question is often discouraged through intimidating "flames" from the more experienced users. Referral to the frequently asked questions (FAQs) is not much help because these are typically not well organized and may be heavily laced with jargon which is unfamiliar to the novice. In contrast, we felt that it was important for a user posing a question to remain anonymous.

Likewise, a user providing an answer may wish to retain a degree of privacy and decide whether or not to reveal their email address to the person asking the question. The commitment expected of an expert in response to any question is thereby limited to a simple answer (and possibly one follow-up question). At any time in the dialog, either party may elect to reveal their identity and thereby invite the other party to establish a longer dialog outside the system.

The desire to respect privacy can also compromise the security of a system. In the Help-Exchange system, the administrator is given privileged access to the identity for all correspondents so that in the event of clear abuse by any user, their account may be closed. (see Security below)

The Question/Answer Dialog

A user posing a question selects the category in the taxonomy that, to the best of their understanding, best fits the topic and for which an expert or experts exist. The question is then posted to the system and assigned to one of the experts in that category. An algorithm is used to conserve "wealth" in the system, so the question will probably be posted to the expert who owes the most answers, for example.

The next time that expert logs into the system, they will see that a question is in their mailbox. Of course, some significant time might elapse before a user logs in, so an email message is sent to notify an expert that a question has been posted to them.

Once the expert reads the question, an answer may be composed and sent, or if the expert does not know the answer, the question may be referred to another expert, either in this domain or another chosen by the expert. If the question is referred, appropriate email messages are sent notifying both the person posing the question and the newly selected expert. The system keeps track of referrals so that the same question does not cycle back to an expert who has already seen it.

Once the question is answered, an email message is sent to the user who posed the question to notify the user that the answer is in their mailbox. If the user feels that the answer is not sufficient, the user may ask one follow-up question which is routed back to the same expert, again ensuring anonymity. Since some context has now been established for the question/answer pair, no referrals are permitted at this point.

Time-outs

The use of email is slightly problematical since it resorts to a second parallel communication system which also retains state. Once sent, an email message may not be retrieved! The expert might be on vacation, or simply busy with other things. Since the user posing the question expects a response in a reasonable amount of time, a reminder is sent to the expert after one time-out and if there is still no response after a second time-out, the question is referred to another expert. If there is no other expert available in the current taxonomy, the question will be posted to the Unsolved Mysteries, a special category described below. Of course, the reminder email messages cannot be retrieved and this may be momentarily confusing to an expert. This problem will hopefully be solved with better integration of email and the web browsers, but is outside the scope of this project.

In addition, the system is self-purging in that each user signs up for a specific period of time. When that time is nearly elapsed, the user is sent an email message prompting them to refresh their account. A reasonable amount of time is allowed for them to do so and if no response is received, the account is deleted. A user, of course, may initiate a new account at any time, but any accrued question/answer credit balance is lost.

Archive

The basic concept of a FAQ archive is useful as long as a reasonable mechanism is provided to assure privacy. In the Help-Exchange system, both the person posing the question and the expert answering the question have to agree to retain their dialog in the archive. At each stage in the dialog, each person may elect to grant permission to archive the question and answer. This continues into the follow-up question and answer and both have to agree to archive the entire dialog. If both agree, the dialog is retained for that topic in the taxonomy. Subsequent users may browse the archive if they wish before posing a new question.

In addition, a user may elect to obtain a copy of the dialog via email for their personal use independent of whether or not the dialog is archived.

Unsolved Mysteries

Occasionally a question may be asked that is not answered by any expert. If an expert to whom the question has been assigned cannot determine how to reclassify the question and if all experts in that category have been exhausted, the question is posted to "Unsolved Mysteries." Any expert may elect to answer a question in this category and receive bonus credit to ask an additional question. This mechanism serves as a last attempt to provide an answer to any question which could not be otherwise answered.

Quality Issues

There are clearly an abundance of quality issues related to use of the Help-Exchange system. Since the experts are self-proclaimed, the answers can only be as good as any individual expert's opinion. The assumption of good will does not assure quality. A supposed expert may be well-intentioned, but still give bad advice. Newsgroups, by contrast, are somewhat self-correcting in this sense since anyone monitoring the dialog (and one cannot help but monitor the dialog if one is a member of a newsgroup) can correct advice that is in error. We do not see an easy solution to this problem (see Future Direction below). The user can, of course, resubmit the question if unsatisfied with the answer including the follow-up dialog. It is likely that the question would then go to another expert, but even that cannot be guaranteed. In its present form, the user of the system can only assume that the answer cannot be guaranteed to be correct and, if really critical, should be independently verified. Still, we assume that users of this system will act responsibly, not only in answering questions, but also in accepting the validity of the answer given the source.

Security Issues

A much more serious question is raised involving malicious mischief. At least initially, the system assumes responsible behavior by its users. Consequently, few security issues have been considered apart from restricting access to the system. Newsgroups and other similar systems are susceptible to "spamming" which is the use of malicious or abusive language which can disrupt or even destroy a system that is unprepared for it. The corresponding waves of protest from innocent recipients of the messages can cause many users to unsubscribe from the group in frustration with the volume of unproductive messages. The Help-Exchange system avoids spamming since the question or answer dialog only involves two people. At any point, any user may complain to the administrator and if a user is clearly abusing the system, they can lose their account privileges. This is thin protection against a determined prankster, but the cascading effect in newsgroups or other similar systems is avoided. Once again, the emphasis is on simple and workable solutions in contrast to what would be required to build a bullet-proof system.

Test Results

The Help-Exchange system has been through two levels of testing so far. After an initial demonstration at the 1995 TCAD Symposium [Losleben & Boning 95], a two week alpha test was performed in late Aug. 1995 with a group of approximately 15 volunteers. The alpha test uncovered a number of design flaws and suggestions for improvements. In addition to the bugs one expects to uncover in testing with real users, a number of useful suggestions were made by the alpha testers. The taxonomy design was completely reworked to allow selective navigation of large taxonomies. Question redirection by an expert was added, as was the Unsolved Mysteries, Follow-up Questions, better navigation aids and the Archive. The most serious bugs had to do with the unique constraints that the web interface imposes on the system design. In particular, web browsers retain state in the form of previously visited pages. Since we rely on the data base at the server side, it is easy for stale pages at the client side to be in conflict with the state of the data base. Unless carefully designed, this produces an awkward or even corrupted interface with the database.

Subsequent to these changes, a longer beta test was undertaken in the Fall and Winter quarters with a larger group of about 30 users at Stanford and MIT. This was the first attempt to satisfy a real need for assistance between users. Overall the beta test ran smoothly and a few improvements were incorporated to make the interface more intuitive and to eliminate aspects of the interface which were error-prone for inexperienced users. More error testing and handling was added and an administrator interface was added to assist in that function. A surprising result of the beta test was the number of new options which were suggested. Many of the users of the system wanted more flexibility in use of the economic model. For example, some would be willing to "pay" more for faster service or to send a question to more than one expert. These suggestions are being considered in balance with our desire to build a simple system.

At this time, a second beta test is planned to begin in Apr. 1996 and is expected to involve about 400 users. Subsequent to that, the system will be made available to a much larger research and education community.

Future Direction

Two major enhancements to the system are being considered. First, in consideration of quality issues, we plan to add a special class of user who is copied on all correspondence within a category of the taxonomy. This is specifically to address a potential application to customer support systems where an individual is assigned the responsibility for monitoring correspondence and correcting problems. This brings the Help-Exchange application closer to systems that support user groups, yet still keeps the volume of traffic down to a minimum for most users.

The most challenging future direction involves the conversion of the system from a centralized system (presently implemented at Stanford) to a distributed system for which independent taxonomies might be developed at other sites, possibly reflecting the unique areas of expertise at those sites. The objective would be to allow a user seeking information to follow a hyperlink to another taxonomy at another site.

This presents two problems, one fairly simple, the other distressingly difficult. First, there would have to be some mechanism for exchange of barter value between sites. There is similar work that is well developed for real monetary systems that the implementation of this is not expected to be difficult. The harder problem is found in accommodating what would certainly be dissimilar taxonomies. One could not expect to enforce a common ordering discipline for taxonomies across a wide range of knowledge domains, or even common ordering disciplines for the same domain developed independently at different sites. For the World Wide Web, an absence of organization is expected, but the implication of a taxonomic organization where that is not likely to happen between sites might be quite confusing. The research for the summer of 1996 is focused on this problem.

Conclusions

The Help-Exchange system is an experimental system intended to provide an alternative to other approaches such as newsgroups for networking researchers and educators who wish to support each other by each offering to provide assistance in their unique area of expertise. The system is web-based and utilizes a taxonomy of disciplines to help connect users with questions to other users who might provide answers. The primary advantage of the system is the greatly reduced email traffic which is often a deterrent to use of other approaches. This paper has described some of the unique features of this system including experience in initial alpha and beta testing.

References

[Borriello 91] Borriello, G. (1991). Microelectronic System Education Conference & Exposition, 1991, CMS Archives, Menlo Park, CA.

[Kantor & Lapsley 86] Kantor, B. & Lapsley, P. (1986)."Network News Transfer Protocol," RFC977, 1986

[Horton & Adams 87] Horton, M. & Adams, R. (1987). "Standard for Interchange of USENET Messages," RFC1036, 1987.

[Losleben 96] Losleben, P. (1996) http://snf.stanford.edu/ComputationalPrototyping/comproto/nre.html

[Gruber 93] Gruber, T.R. (1993) "Toward principles for the design of ontologies used for knowledge sharing." In Formal Ontology in Conceptual Analysis and Knowledge Representation, Nicola Guarino and Roberto Poli, editors, Kluwer Academic, in preparation. Original paper presented at the International Workshop on Formal Ontology, March 1993. Available as Stanford Knowledge Systems Laboratory Report KSL-93-04.

[Losleben & Boning 95] Losleben, P. & Boning, D. (1995) "A New Research Paradigm using Internet Collaboration or Building a National Research Enterprise," TCAD Symposium, Hierarchical Technology CAD--Process, Device and Circuits, Aug. 1995, Staford, CA.

Acknowledgments

The authors gratefully acknowledge the support of the Defense Advanced Research Projects Agency under Contract DABT63-94-C-0055.

From a Set of Technical Documents to a Hypertext System on the Web

Yannick Marchand, Jean-Luc Guérin, Jean-Paul Barthès
Université de Technologie de Compiègne
U.R.A. C.N.R.S. N° 817 HeuDiaSyC
B.P. 529 - 60205 Compiègne Cedex - France
E-mail : {marchand,jlguerin,barthes}@hds.utc.fr

Abstract: Mosaic and Netscape have democratised Internet and popularised the notion of hypertext. Although we are clearly satisfied in having access to large amounts of information, we can, on the contrary, regret that at the same time the concept of hypertext has been simplified to such an extent that it has eclipsed some of the original defined goals and hopes of its creators. This article has two objectives. Firstly, to exhibit how this departure from the creator's ideas occurred. We will therefore take a look at the origins of hypertext by considering the aspirations of its creators, namely Vannevar Bush, Douglas Engelbart, Theodor Nelson, and also Paul Otlet, an often overlooked visionnary author. Secondly, nevertheless, we will illustrate how it is possible to create a 'true' hypertext on the Web. The hypertext prototype (entitled 'Nestor') that we present has been developed for France Télécom and is notably based upon the coupling of an object oriented database and Netscape through the intermediary of a script language.

Introduction

Initially conceived in the U.S.A. in 1969, during the cold war, for the benefit of a military preoccupied by an eventual rupture of communications, the Internet was immediately used by research bodies and universities, in order to exchange their ideas. It was with this perspective in mind that the World Wide Web was created, to support a hypermedia dedicated to building information servers on the Internet. Created in March 1989, on the initiative of Tim Berners-Lee [Berners-Lee 94] for the community of physicians at the Centre d'Etudes et de Recherches Nucléaires (CERN) in Geneva, its aim was to centralise scientific results, publications and documentation. The Web was destined to be a great success and was transformed into a cultural and social phenomena when, in February 1993, Marc Andreessen of the National Center for Super Computing, University of the Illinois, U.S.A., edited the first version of NCSA Mosaic [NCSA 93]. This graphical interface and Netscape will greatly accelerate the development of the Internet. The Internet already has millions of users and will soon be affordable to the average household and, and as a direct consequence, it is attracting the interest of companies throughout the world.

As the Web has been completely successfull, due to the fact that it retrieves information in a quick, powerful and intuitive manner, the approach and the technology used, especially the hypertext, will become well known. Unfortunately though, the concept of hypertext has been simplified. In order to understand how this concept could lose some of its original meaning, it is necessary to take a fresh look at the precursers of hypertext and describe their motivating factors. This will facilitate the presentation of 'Nestor', a prototype hypertext for the Web that we have developed for CNET[1] at Lannion, one of the main research centers of France Télécom.

The Founding Principles of Hypertext

The names of Vannevar Bush, Douglas Engelbart, and Theodor Nelson invariably get mentioned when the recent history of hypertext is under discussion. Indeed, the projects Memex [Bush 45], Augment [Engelbart 68], and Xanadu [Nelson 88], that are expressions of new ideas or concrete realisations, have been crucial to the development of this research domain.

It is necessary to lament the absence of Paul Otlet from the aforementioned group of scientists because this Belgian author exhibited in his work [Otlet 34], 11 years before Bush, an exceptional clairvoyance bordering on prophecy . It is for this reason that this section uses the ideas of these four pioniers of hypertext to exhibit three essential elements that characterise and justify the fact that the word 'hypertext' means etymologicaly 'more than' 'text'.

The Gift of Ubiquity

Noting that the number of books and documents increases every day, [Otlet 34] proposes, in order to confront this deluge of information, the creation of 'bibliology', a science and general technique for documentation. The creation of this science would necessitate "a set of interlinked machines" having to perform seven operations of which "the establishment of documents in such a manner that each piece of data has its own individuality and in its relations with other data, it must be called anywhere that it is required" (operation 3) and "automatic access to consulted documents" (operation 6). It is obviously possible to note that these four authors shared the same preoccupation : the organisation of literature on a large scale, support for knowledge accumulated across the centuries, in order to make access easy and quick to that which is being manipulated. Within the projects Memex and Augment, the aim is to help the researchers with their reseach documents. The aim of the Xanadu project is slightly different, its aim being the construction of an immense network that takes into account all the available documentation ever published.

It is possible that these human and ambitious aspirations are themselves cemented within the Web. Indeed, the Web plays the role of a global library, giving its users a flexible and immediate access to a set of documents that are spread worldwide. Thus, the Web gives the impression that the user is consulting a unique document although, in reality, the user is visiting several separated servers throughout the world. Due to the dematerialisation of documents and abolishment of the notions of distance and time, the Web offers amazing possibilities, by using simple electronic clicking, to be everywhere at once. It should be noted that Engelbart, by the invention of the 'mouse' and experimentation with multi-windowed screens, has greatly contributed to an instantaneous and associative displacement within the jungle of information. In other terms, this displacement inspires, as wished for by [Bush 45], our natural manner of thinking ("As We May Think").

The Omnipotence

[Otlet 34] mentions a second essential principle for the concept of hypertext. It concerns the "presentation of documents, either by viewing directly or through the intermediary of a machine that has to make additional inscriptions" (operation 6) and the "mecanical manipulation, at will, of all the recorded data, in order to obtain new combinations of facts, and new relationships between ideas" (operation 7). Within this outlook, the user is no longer only passive, content to consult elements of information connected by active links, but active as well, in the sense that the user has available these elements in order to add annotations and personal links to them. This is the reason why the boundary between the author and reader has a tendancy to disappear since the reader benefits from a freedom comparable to that of a sculptor who is allowed to model, at will, using the material that is initially given.

On the Web, a non computer-literate user unfortunately can not exercise this freedom of action on the documents. Indeed, the creation of one link for the user, for example, is neither natural nor convivial because it is necessary to have minimal knowledge of the following : (i) directories and files, (ii) text editors, and above all (iii) the language HTML (HyperText Markup Language) [Morris 95] that is used to describe the documents. This creation of the relation can be considered as an important intellectual act since it constitutes, for its author, an argumentative and rhetorical element.

The Omniscience

"The machine that would perform these seven operations would be a veritable mechanical and collective brain"[Otlet 34]. "An active community will be constantly involved in discussion concerning the contents of its manual"(Engelbart). These two quotations put the accent on the last distinctive characteristic of the concept of hypertext, namely the cooperative work that puts the creation of personalised links and commentaries within the social construction of knowledge. Due the fact that a hypertext is adaptable and shareable, this approach means that it is never a final product but remains, for its users, an area of expression and memory that is constantly evolving. The hypertext therefore takes the form of a flexible tool of social communication, at the service of collective intelligence processes [Lévy 90]. Thus it becomes possible for each user to have access to all of the knowledge acquired by the community. At the time of his writing, Bush could already imagine a new profession of trail blazer who would be the type of experts capable of discovering and building useful routes within these documents.

It is certainly this characteristic that illustrates the most the difference between the Web and the first aspirations of the concept of hypertext. Due to the fact that the Web is organised according to a client-server architecture, each author is only in charge of a limited number of documents, of which the author has sole rights to define the links to other documents. In other terms, the documents that have not been created by the author are consultable but communication itself does not exist, since it is not possible for the user to adjust and transform them. In this case, it consists more of an interconnection of distributed knowledge : each user puts his knowledge at the disposal of the collective and knows that he can access, by return, all the information that he requires but does not have in his possession [Nanard 95]. "I offer to others my microcosm of documents" has substituted the original idea of "Let's share the universe of documents that we transform together".

Presentation of 'Nestor'

It is possible to rename technical specifications as documents that define the characteristics of a product or service. These specifications have to comply to certain recommendations (or standards), namely a set of rules that are normally created by international organisations of standardisation. This section describes how as such recommendations have been treated with respect to 'Nestor', a hypertext prototype that has been developped for CNET at Lannion.

Characteristics of Corpus and Objectives

For the personnel who have to write specifications, the corpus of recommendations can look like an encyclopedia. Indeed, these reference documents give, in the form of english text, information concerning, amongst other things, definitions, concepts, and examples. The recommendations form a "microcosm" of interdependant documents and are structured in the form of traditional linear texts, namely with a contents page, and a set of successive paragraphs, grouped in chapters. Containing multiple internal and external references, specifically to other documents, the consultation of these 'spaghetti documents' is based as much upon a mechanism of the association of ideas as it is upon a sequential and chronological reading.

The aim of 'Nestor' is to transform the set of recommandations into a hypertext for the Web. It is worthwhile asking if this transformation is opportune, because, as stated by [Nielsen 90], "just as the best films are not made by putting a camera in the front row of a theater, the best hypertexts are not made from text that was originally written for the linear medium". In response to this objection it is possible to put forward, in our case, the following two arguments :

- From a practical point of view, it would be laborious to restructure these texts. The people responsible for documenting the Ariane space project state, "A rocket produces its own weight in paper". Within the domain of telecommunications, documentation is both voluminous and complex.
- From a methodological point of view, this transformation seems to be justified, since it is based upon the conviction that each document adds value to others, that the whole is greater than the sum of the parts [Glushko 89].

Due to the fact that 'Nestor' is a hypertext prototype, it should logically display the characteristics that have been already stated for the gift of ubiquity, the omnipotence, and the omniscience. These terms were voluntarily emphatic in order to emphasize the quasi-divine character of the concept of hypertext, to enable it to be initially defined. In our case, it concerns recalling the last two founding principles, which happen to be the most commonplace terminology, of easy creation of personalised links and team work.

Nodes and Typology of Links

The hypertexts have to, as their first task, articulate and organise the entities of information (nodes), by use of relations (links) that exist between these grains of knowledge. These links are activated by the user in order to travel elsewhere, according to his interests.

Within 'Nestor', the nodes represent the formal and logical divisions that can be found in reference documents, specifically, the contents page, the chapters and optionally the appendices. Coming from this 'natural' division of the units of meaning, it is possible to define the following links :

- 'Zoom'links (or focalisation links) : these join the contents to chapters.
- Plan links : links between chapters and the corresponding contents (these are the inverse of 'zoom' links).
- Succession links : links used for linear reading, as in the case of a document (this is the notion of following chapter).
- Precedence links : inverse of succession links (this is the notion of previous chapter).
- Reference links : these join a word (or a group of words) of a chapter to either a contents page or to another chapter. The references can be external and internal. They are usually preceded by an expression such as "cf.", "see", "for example" etc.

These links, that could be renamed as structural links due to the fact that they are directly derived from logical organisation of linear texts, can therefore be automatically identified and generated by a compiler. In additon to these objective links, there are subjective links [Kahn 89] that are reference links created by the users in order to enrich the initial connectivity of the hypertext. This category of personalised link is important for hypertexts as it gives the possibility of adding links that transform, in a certain way, each reader into a potential author. Thus the user can, in concrete terms, structure and build his own knowledge. Indeed, as time progresses, the user, by interacting with the hypertext, acquires an understanding of the domain under exploration [Yankelovich 87]. The hypertext therefore becomes, for the user, a depositary of expertise, a way to organise his knowledge by correcting and completing, through the creation of personalised links, the incoherencies and deficiencies of initial texts. It should be noted that, in our case, only the addition of links can be performed, since the number of nodes is, by definition, constant as it is determined by a finalised set of documents.

Therefore, it is possible to say that, in the first instance, the user appropriates the knowledge of texts through the use of structural links. It is only later that he reappropriates this knowledge through the use of personalised links. In order to explain the semantics associated with a personalised link, the user can add 'commentary' details to augment the anchor and the node of destination. Due to the fact that the 'commentary' and 'destination' details are optional, the personalised links found within 'Nestor' can be classified as one of the following three types :

	Destination	Commentary
Direct links (DL)	Yes	No
Commented links (CL)	Yes	Yes
Simple commentary (SC)	No	Yes

In the case of the link with commentary, the term 'link' is used even though it is not a real link, due to the fact it does not relate to two nodes of the hypertext but to a node that is connected to a commentary page (this is the reason why the acronym 'SC' signifies 'Simple commentary'). A personalised link is public or private. The public link, in opposition to the private link, means that the author's work can be accessed by other users.

The General Architecture

The Web uses the CGI (Common Gateway Interface) [Mc Cool 94] that serves the purpose of writing the bridges between the information HTTP servers (HyperText Transfer Protocol) and external programs. The role of these programs, that are commonly known as scripts, consists of : (i) capturing the parameters entered by the user, (ii) manipulating them and, (iii) giving a result to the client program that made the request. The architecture chosen for 'Nestor' is based upon the coupling of Matisse[2], an objet oriented database, and Netscape, a client of the Web. The interfacing of these two applications is assured by scripts written in Python, an objet oriented programming language developed at the 'Centrum voor Wiskunde en Informatica' (CWI) of Amsterdam [Van Rossum 93].Matisse is a system used for the managment of object oriented databases. Its basic concept is the PDM model (Property Driven Model), developed at the University of Technology of Compiègne [Barthès et al. 86]. This model is based upon both semantic networks [Quillian 68] and frames [Minsky 74]. An object is characterised by two different properties : the attributes and the relations. The notions of minimal and maximal cardinality are associated with the relations, as occurs within the Entity-Association data model [Chen 76]. From a practical point of view, it is possible to note the following :

- all values added to the documents are systematically assigned a signature that comprises the date at which these improvements have been performed, and the name and electronic address of the author.
- the creation of a personalised link is both simple and dynamic. In the first instance, the author selects the line, and then the word, that will play the role of anchor. Secondly, the 'Destination' and 'Commentary' fields enable him to classify the links (cf. the end of section 3.2).
- the equality approach, that consists namely of all the users having access to operations in order to improve documents, is only verified for the members of CNET. This explains the requirement for the use of personal passwords within Matisse, in order to identify users. Whilst using 'Nestor' with a password that has the specific value 'anonymous', an external CNET user would be able to, nevertheless, consult the recommendations and all the additions before they are declared public.

Python is an interpreted programming language that proposes objects and high level operations using a simple syntax that is based upon indentation. In addition, it has the advantage of possessing a standard 'CGI' module that allows the easy capture of parameters from a HTML page. The following modules have been written for 'Nestor' :

Matisse

Within this module are the main functions (written in C) of the Application Programming Interface of Matisse.

FormatHtml

This module embodies a part of the HTML language. This language is used for the diffusion of documents by Web servers and consists of a set of formatting commands.

Compiler

The compiler allows the transformation of the set of ASCII recommendations into HTML files. This transformation is performed in two stages. Firstly, the compiler identifies the units (or nodes) of the hypertext, more specifically the contents page, the chapters, and the appendices, and gives them an identifier (number of nodes) that will be strored within Matisse via use of the 'Matisse' module. Secondly, the compiler segments these units in order to physically generate HTML files by use of the 'FormatHtml' module. This generation constructs the structural links and prepares, for each node, the options that will permit improvement (cf 'Personnalisation' module). It is easy to understand why, for reasons of security, the 'Nestor' option that uses the compiler is safeguarded by a password.

Personnalisation

This module offers the possibility to the users of improving the recommendations by the intermediary of annotations, key words, and especially personalised links. In the latter case, the user can have, at any moment for the current node, a snapshot of all the public personalised links that the group possesses. In the same manner, the plurality of viewing points on a anchor must be taken account. For instance, the following expression : **Anchor** (Cl) *(Sc) (Dl)* indicates to the user that :

- three links are associated to the anchor, namely a commented link, a simple commentary, and direct link,
- the last two links were not created by the user, indicated by italics,
- the signature of all these links and their complementary information is accessible by the activation of the symbol.

Conclusion and Perspectives

The 'Nestor' prototype has been recently experimented in a real environment. Two types of ASCII documents were compiled, namely, a set of recommendations for CNET and documentation dedicated to the Python programming language. The results of the two compilations are given in the following table :

	Documentation Python	Recommendations CNET
Number of documents	2	6
Number of bytes	183 471	518 845
Number of pages	86	213
Number of nodes	122	85
Number of structural links	427	408
Number of reference links	11	104

We are currently looking for people willing to test 'Nestor' in order to gather comments about its running and its ergonomics. The adaptability manifests by the fact that the user enriches the hypertext via his own knowledge. In the notion of adaptivity, the hypertext system takes the initiative by proposing links that might be relevant to the user. In order to find such informal links, a neural approach, based upon the Hopfield model, is currently under development.

Literature References

[Barthès et al. 86] Barthès, J-P., Vayssade, M., & Monika Miaczynska-Znamierowska (1986). Property Driven Data Bases. Internal rapport of the University of Technology of Compiègne.

[Berners-Lee 94] Berners-Lee, T. (1984). World Wide Web Initiative, CERN - European Particle Physics Laboratory. http://info.cern.ch/hypertext/WWW/TheProject.html

[Bush 45] Bush, V. (1945). As we may think. Atlantic Monthly, July 1945, 101-108.

[Chen 76] Chen, P.P. (1976). The Entity-Relationship model : towards a a unified view of data. ACM Trans. on Database Systems, 1 (1), 9-36.

[Engelbart 68] Engelbart, D. (1968). A research Center for Augmenting human intellect. AFIPS conference proceedings, 33 (1). Thompson books, Washington DC, 1968.

[Glushko 89] Glushko, R. (1989). Design Issues for Multi-Document Hypertexts. Proceedings Hypertext '89, Special Issue SIGCHI Bulletin, 1989, 51-60.

[Kahn 89] Kahn, P. (1989). Linking together books : Experiments in adapting published material into hypertext. Hypermedia, 1 (2), 1991, 111-145.

[Lévy 90] Lévy, P. (1990). Les technologies de l'intelligence. L'avenir de la pensée à l'ère informatique.

La Découverte.

[Mc Cool 94] Mc Cool R. (1994). The Common Gateway Interface. National Center for Supercomputer Applications, University of Illinois at Urbana-Champaign, http://hoohoo.ncsa.uiuc.edu/cgi

[Minsky 74] Minsky, M. (1974). A framework for representing knowledge. Cambridge Mass., M.I.T.

[Morris 95] Morris, M. (1995). HTML for fun and profit. SunSoft Press, Prentice Hall.

[Nanard 95] Nanard, M. (1995). Les hypertextes : au-delà des liens, la connaissance. Sciences et techniques éducatives,. 2 (1), 1995, 31-59.

[NCSA 93] National Center for Supercomputing Applications (1993), University of Illinois at Urbana-Champaign. NCSA Mosaic. A WWW Browser, http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/fill-out-forms/overview.html

[Nelson 88] Nelson, T. (1988). Managing Immense Storage : Project Xananu provides a model for the possible future of mass strorage. Byte, 13 (1), January 1988, 225-238.

[Nielsen 90] Nielsen, J. (1990). Hypertext and Hypermedia. Academic Press.

[Otlet 34] Otlet, P. (1934) : Extracts from "Traité de documentation, le livre sur le livre" taken from "La Pensée", n°281, May/June 1991, 66-71.

[Quillian 68] Quillian, R. (1968). Semantic memory in Minsky, M. (ed.). Semantic information processing, Cambridge Mass., M.I.T. Press, 227-270.

[Van Rossum 93] Van Rossum, G. (1993). An Introduction to Python for UNIX/C Programmers, in the proceedings of the NLUUG najaarsconferentie 1993.

[Yankelovich 87] Yankelovich, N. (1987). Creating Hypermedia Material for English Students, Sigcue-Outlook.

Acknowledgements

Special thanks to (i) CNET, for funding the project, (ii) Jean-Pierre Poitou (CREPCO, University of Provence), the first person who advocated Paul Otlet 's writing to us and (iii) Darren Millward, for his precious help during the compilation of this English version.

Catalog: a WWW Gateway for DBMSs

José M. Martínez, Francisco Morán
Grupo de Tratamiento de Imágenes, Depto. Señales, Sistemas y Radiocomunicaciones
E.T.S. Ingenieros de Telecomunicación, Universidad Politécnica de Madrid
28040 Madrid, Spain
Tel: +34.1.336.7353
Fax: +34.1.336.7350
Email: {jms,fmb}@gti.ssr.upm.es
WWW: http://www.gti.ssr.upm.es./staff/{jms,fmb}UK.html

Abstract: The Catalog software package described here is designed to act as a gateway to the WWW for information systems based upon a DBMS: a so-called *hyperbase* program. There exist already several approaches to the problem, mainly working with ORACLE's RDBMS, although migration to most of other vendors' RDBMSs should be straightforward. Catalog handles the problem in a new way, and to serve a different primary purpose: it is a tool to dynamically extract items from the DB and group them together. No modification in the DB structure is required, and it is not necessary -nor possible- to program CGIs for querying. Therefore, its main role is the dynamic creation of *Cards:* HTML pages that are the result of searching and retrieving the contents of the DB.

Introduction

Hyperbases and the WWW

Hyperbases [Hunter et al. 95] are hybrid systems in which information from a relational database is merged into hypertext documents for presentation. The World Wide Web has facilities which make the provision of hyperbases possible. The basic capability of WWW is to download files via hypertext links. Dynamic documents are programs which can be invoked in place of a direct file download, and generate the document to be downloaded as its output; a suitable program can hence extract information from a database and present it as an HTML document. For instance, the Forms interface, available under CGI, allows web pages to include user-interface elements such as fields, buttons and check-boxes: this can be used to provide data input to dynamic documents.

Hyperbase programs may be presented to the user in two ways. First, the user may be presented with forms which ask explicitly for input: for example, key-words to be used in a search. Forms can also be used to present output, although this is rarely done: most hyperbase programs present the results of a search as "simple" hypertext. The main problem with such hyperbases is that they require sophisticated CGI-compliant programs to interpret forms, query appropriate databases, and merge the resulting information into hypertext. There is a clear need for application-generator tools to allow hyperbase programs to be constructed with minimal expertise on the part of the designer. The other possible solution is to let the user browse what appear to be normal web pages (i.e., with no forms), but are in fact being dynamically constructed from the database.

Related Work

The WOW, DECOUX, ORAYWWW and WORA Gateways [Oracle 95] are some examples of tools that were developed before the release of the ORACLE Web Agent included in the ORACLE Web Server [Oracle 96], and go some way towards merging WWW and ORACLE. Another gateway that needs programming, but is open to different RDBMSs is GSQL [Jason 95].

Catalog does not need programming as the previous gateways do. Two other approaches which do not need programming either are SWOOP and MORE, which use ORACLE as a back-end, but are easily portable to other vendors' DBMSs:

- SWOOP [Hunter et al. 95] provides an application generator, namely a CGI programming language and environment. Information is presented to the user by means of simple hypertext pages; these pages are specified using an augmented yet simple HTML syntax. The system designer using SWOOP needs some knowledge of ORACLE and HTML, and the programmer should know about the DBMS schemes.
- MORE [Eichmann et al. 94] provides a meta-data repository: information about the artifacts under its management scope is stored and consulted to locate, load and compose those artifacts, stored by other means. MORE acts as a front end to organize assets based on a predefined meta-data repository, where the assets are incorporated. Browsing can be done by collection, classes, and natural language or pattern searches.

Catalog needs neither ORACLE nor HTML knowledge, as SWOOP does, but on the other hand, there is no flexibility for programming searches in the DB. It also leaks the complex searching facilities of MORE, but no reorganization into a meta-data repository is needed.

A closer approach to Catalog's is WDB [Rasmussen 95], which provides a software tool-set that simplifies the integration of SQL databases into WWW, providing access to the contents without writing a single line of code. The interface consists of a WDB script written in Perl and a set of *Form Definition Files (FDFs)*, each describing a different view of the database. These FDFs play a role similar to the Catalog *Templates*. Although Catalog does not include currently data conversion and computed values of fields, it provides access control, multilingual support, Card layout customization, and a more elaborated query mechanism.

Catalog system

The principal aim of Catalog is to provide an immediate and easy availability of data collections, as already available at several sites (museums, libraries, etc.). Besides the dynamic creation of queries, sets of results (Card descriptors), and Cards, it offers the means for:

- Customization of the Card layout (by the modification of the Templates).
- Dynamic information filtering over the data, thus allowing the retrieval of different parts of the data depending on the user profiles defined by the Information Provider - and independently of other access control mechanisms provided by the Service Provider.
- Supporting multilingual dynamic page creation.

Catalog consists of a *Templator* creating the raw templates of the user's tables in the database, and a set of compiled CGIs providing the set of services needed to search, retrieve and navigate through this information. The customization of the templates makes it possible to use fully all the available features; it follows a set of simple guidelines where no programming is required, nor DB or HTML knowledge. Every Catalog *Service* is divided into three software blocks [Fig. 1]:

- The *CGI interface* for that particular service, which interprets the input parameters (POST and GET methods), calls the needed functions, and formats the result into an HTML document.
- The *Functional Core*, composed by a set of functions for searching and retrieving the data items. These functions use the Database Interface for retrieving data from the storage system, and the user's Template for information filtering and object composition (including multilingual labelling).
- The *Database Inteface*, which allows the access to the available storage system, be it a DBMS, a file system, or any other. It is composed by a couple of simple C functions. Currently, it is available only for ORACLE, and it uses the Pro*C precompiler.

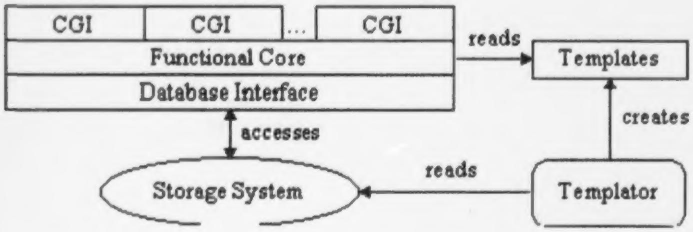


Figure 1: Catalog software blocks

Catalog services

There are several Catalog services, each with a separate CGI interface, to allow a flexible control over data retrieval, although not all of them are needed for all access configurations.

entries Reads the login and password for a user introduced via an HTML form. The dynamic document includes the available entries for this user.

fields Returns, for a given user and an entry, a dynamic document containing the available search fields. Several fields can be chosen from the list.

items Returns, for a set of search fields, a dynamic document containing the list of available items for each fields (if the field is of type WRD - see below), a free search box (types FRE or TXT), or both (type STR). Several items can be chosen for creating a query. The items of a same field are grouped by an OR condition, while different field groups are grouped by an AND condition.

hits Returns, given a query, the number of hits and a list with the *Card's labels*. If there are less than a pre-defined number of hits, the dynamic document contains a set of descriptors instead of the list of labels. Several items can be selected from the list of labels in order to ask for the descriptors.

shorts Returns a set of descriptors for a list of Card's labels. Each descriptor holds a link to the *full Card.*

card Returns a Card.

In order to create the dynamic document, each CGI service first reads the Template for information filtering and document composition and labelling instructions, and then retrieves data from the storage system. This is done using the Functional Core services.

Functional Core

The Functional Core provides the general functionalities needed by the particular services. For each CGI there is a corresponding Functional Core public service, which is in turn composed by one or several private methods in order to perform information filtering, access control, data locations and data retrieval. The formatting of the document into HTML is done within the public service code, in order to allow future reusability of the private services for other information servers supporting different document formats like MHEG.

Once it has read the Template and retrieved the document components, the Functional Core checks the permission level for each component with the user access level. When the Functional Core has identified the components that are accessible for that particular user, it accesses the Storage System through the Database Interface to retrieve the data, and then composes the document according to the Template structure.

Currently, access control is implemented by means of a simple, file based, security mechanism. When the user logs in the system through the entries CGI, the security agent checks the identity and finds the user profile associated to this user, which corresponds to a per site access level. This access level is used for information filtering using the permission levels associated to the different entries and fields via the Template customization. The user profile is the base for a *session ticket*, which includes information about the session language. Both values are passed among CGI and dynamic documents during the session, in order to obtain session state information. For security, the session ticket is encrypted with a key which varies from session to session.

Database Interface

In order to allow independence from the underlying DBMS, the Functional Core translates all the queries into a set of simple SQL calls which can be easily implemented for every SQL-DBMS. The required functions are: **connect_db, disconnect_db, count_hits, get_hits**.

For the moment, only the ORACLE Database Interface has been implemented using the Pro*C precompiler. INFORMIX and Ingres versions are planned as well, using the corresponding C precompilers. Some work is also planned for implementing this interface over a propietary DBMS not supporting SQL, but only keyword search. This will yield to a redefinition of the current interface, which is clearly RDBMS-compliant.

Catalog Templates

Catalog uses templates for both information filtering and document composition, including multilingual labelling. There must be a Templator application for each supported DBMS, that will create a master template for each DBMS user. This master template extracts all the tables, columns and cross-links for that user, so that the Catalog CGIs can run properly, although it may need customization for a full use of Catalog. For each table that will be accessible to the user, an entry is created in the master template with the following syntax:

```
**
IU_REF,IU_link,permission_level,label
number_of_components
{component}
...
{component}
IU_END,IU_link
**
```

where:

IU_REF Information Unit REFerence begin tag.

IU_link The table's primary key and the table name separated by the '~' character, as in primary_key~table_name.

permission_level An integer assigning a minimun permission level for accessing the table.

label A label for the Information Unit or component. The Templator writes the table name or column name, removing '_' characters, if any. In the case of a link component (REF:IU type - see below), the Templator writes LINK*i*, where *i* is the number of the link in the IU's table.

number_of_components
The number of componentes found for this Information Unit.

{component} component_type,component_link,permission_level,reference_flag, search_flag,document_flag,descriptor_flag,label_flag,label

component_type One of STR, NUM, etc.

component_link If the component corresponds to a table column, the column name; if it corresponds to a link between tables (REF:IU type), the link definition including both anchor (current table's column name) and destination (column and table names), separated by the '~' character, as in column_name~column_name~table_name; if it corresponds to a group tag (LAB or BAL), a group style (see also below).

reference_flag I or R, for inclusion of the component data or only a reference to it.

search_flag SEARCH or NOSEARCH, depending on whether the component is searchable or not.

document_flag DOC or NODOC, depending on whether the component should be included in the Card's document or not.

descriptor_flag SHORT or NOSHORT, depending on whether the component should be included in the Card's descriptor or not.

label_flag LABEL or NOLABEL, depending on whether the component should be included in the Card's label or not.

IU_END Information Unit reference END tag.

component types:

WRD Textual component available for list of terms searches.

FRE Textual data, with free text search.

STR Textual component presenting list of terms and free text searches.

TXT Textual preformatted data, with free text search.

NUM Numerical component, currently not searchable.

DAT Date component, currently not searchable.

BIN General binary data component.

IMA Image component.

VID Video component.

AUD Audio component.

REF:IU Link from a IU's component (entry field) to another IU, created when there are referential constraints between tables.

REF:{BIN,TXT,IMA,VID,AUD}
Link to a binary, textual, image, video, audio component.

URL Internet resource locator component, which is presented as an active HTML link.

LAB Group begin tag.

BAL Group end tag.

group styles:

DEF_LIST The group components are presented as a definition list, with the label begin the definition term and the content the description. This is the default style.

ROW The group components are presented each in a row.

COL The group components are presented each in a column.

NUM_LIST The group components are presented in a table cell, each being an item of a numbered list.

LIST The group components are presented in a table cell, each being an item of a unnumbered list.

The page layout is based on HTML tables, in which each component is an item in a definition list consisting of a unique cell table (the document and descriptor Cards have some minor format differences). This is the default format style, which can be modified by using the group tags, although the table is still the base. Of course, the groups are recursive, that is, they can be combined in order to compose complex Card layouts.

The Templator converts the ORACLE types into the following Catalog types:

- VARCHAR2, CHAR STR
- LONG TXT
- NUMBER NUM
- DATE DAT
- RAW, LOW RAW BIN
- Referential constraints between tables REF:IU

Via customization, it is possible, for instance, to:

- Reorder the components within the IU.
- Group components and add grouping labels (via the label field of the LAB type component).
- Customize layout using LAB components with group styles in the component_link field.
- Identify BIN types as images, video, audio or other.
- Indicate that STR components are references (e.g., an image or video file path, that is, REF:IMA or REF:VID).
- Set permissions to the desired values.
- Change the labels to "natural language" ones.
- Add multilingual support. The current implementation uses the component's label, where several language labels are separated by the '%' character. The order of the different language labels, should correspond to the ones defined in the login form or in any other first form. The first one is used as default one.

Installation and Availability

To install Catalog, it suffices to copy the CGIs to the directory where the CGI binaries of the Web Server reside, and to run the Templator in order to create raw templates that can be eventually modified later on.

Currently, a prototype version of Catalog is available for consulting the database of our research group at "http://www.gti.ssr.upm.es./", and also that of the Museo Arqueológico Nacional (Spanish National Archaeological Museum) at "http://www.gti.ssr.upm.es./~man/".

Conclusions and Future work

The purpose of Catalog is the dynamic creation of HTML pages reflecting the search and retrieval of the contents of the DBMS, that can be interpreted as Cards. This has been fulfilled, with no need of modifying the storage structures already established by the data providers. It also allows information filtering and multilingual labelling.

The principal enhancement should focus on the searching mechanism, to avoid the "rigidity" of the present queries. A more friendly tool to edit templates would also be helpful, as well as a more powerful layout definition support. The access control is being enhanced by using a more elaborated security mechanism.

References

[Eichmann et al. 94] D. Eichmann, T. McGregor, and D. Danle. *Integrating Structured Databases into the Web: the MORE System.* Computer Networks and ISDN Systems 27 (1994) 281-288.

http://rbse.jsc.nasa.gov./eichmann/www94/MORE/MORE.html

[Hunter et al. 95] A. Hunter, R.I. Ferguson, and S. Hedges. *SWOOP: an application generator for ORACLE/WWW systems.* http://www.w3.org./pub/Conference/WWW4/Papers/207/

[Jason 95] Jason Ng. *GSQL - a Mosaic-SQL Gateway.*
http://base.ncsa.uiuc.edu.:1234/gsqlsrc/starthere.html

[Oracle 95] *The Oracle World Wide Web Interface Kit.*

http://dozer.us.oracle.com.:8080/index.html; http://www.oracle.com/corporate/press/html/sdkrel.html

[Oracle 96] *Oracle WebServer: About Oracle Web Agent.*

http://www.oracle.com./products/websystem/webserver/html/ois3.html

[Rasmussen 95] B.F. Rasmussen. *WDB - a Web Interface to SQL Databases.*

http://arch-http.hq.eso.org./bfrasmus/wdb/intro.html

Browsing Hypermedia Composites: An Algebraic Approach

Hermann Maurer, Keith Andrews and Nick Scherbakov
Institute for Information Processing and Computer Supported New Media,
Graz University of Technology, Schießstattgasse 4a,
A-8010 Austria.
Tel.: 43/316/83-25-51/Ext. 12
Fax: 43/316/82-43-94
Email: nsherbak@iicm.tu-graz.ac.at

Abstract: This paper presents an algebraic approach to the browsing of hypermedia composites, based on a fully object-oriented, logical data model for hypermedia called the HM Data Model.

The main unit of abstraction in the HM Data Model is an object called an S-collection. Hypermedia links are encapsulated within a particular S-collection and are bound between S-collections. The algebraic operations INCLUDE, JOIN and DIVIDE can be applied to S-collections to temporarily create new views of the hypermedia structure.

Introduction

Hypermedia data models serve as necessary abstractions in order to combine large amounts of raw multimedia information into a logically integrated hypermedia database and provide users with a well-defined set of operations for creating, modifying and accessing such a database.

In fact, the term hypermedia was born from a particular data model called the *node-link* paradigm. In this basic hypermedia paradigm, information is stored as a so-called hyperweb (i.e. as a collection of primitive *nodes* , with *links* between nodes representing different relationships). In this node-link data model, access to information is achieved by freehand browsing of the hyperweb. The *default browsing semantics* introduced by the basic node-link paradigm and then adopted by nearly every logical hypermedia data models can be defined as follows: if there exist a number of computer-navigable links emanating from a particular node, then any link can be manually selected in order to traverse the hyperweb to the destination of this link (i.e. in order to access another node) where the process is repeated.

As hypermedia systems become larger, more dynamic and possibly distributed, it becomes less and less possible to create and maintain the entire hyperweb of information solely by means of the basic node-link hypermedia paradigm

In this paper, we discuss a so-called HM Data Model and offer algebraic operations to extend the navigational component. The rest of the paper is structured as follows: Section 2 contains a brief description of the HM Data Model and an illustrative example. In Section 3 we discuss the algebraic approach to browsing hypermedia composites supported by the HM Data Model. The paper ends with some concluding remarks.

The HM Data Model

Data Structure Types

The HM Data Model defines a particular hypermedia database as a set of abstract data objects called *S-collections* ("structured-collections"). S-collections are created as instances of previously defined system classes.

Every S-collection (instance) has a unique identifier, a name used by other S-collections for addressing, and generally a "data unit" which will be called *content* henceforth. Executing the content will ordinarily cause some information to be presented (text, pictures, audio, video-clips, etc.).

An S-collection encapsulates a particular internal structure. The internal structure is a set of other S-collections (called *members* henceforth) related by a number of computer-navigable links. One of the members of the S-collection must be designated as *head* of the S-collection. Note that links are encapsulated within a particular S-collection: they may only be defined between members of the same S-collection. In this sense, links belong to a particular S-collection; they do *not* belong to the hypermedia database or to either of the members related by the link (hence the "local referential integrity" [11] of the model). Links also do *not* exist as independent data objects which are instances of a particular class, operations cannot be addressed to links.

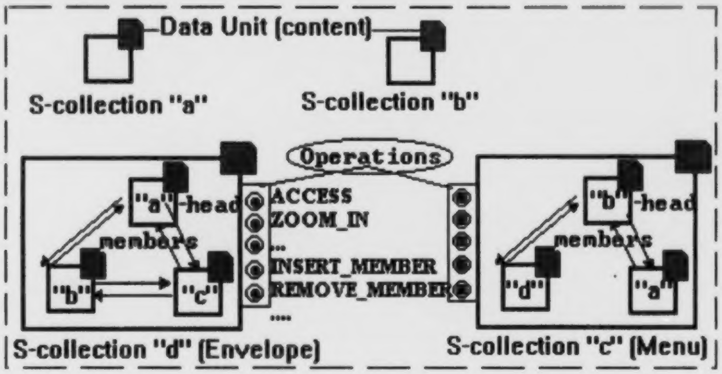


Figure 1: Internal Structure of S-collections

The HM Data Model provides a number of predefined subclasses of S-collection. From a data structuring point of view, these predefined subclasses are similar to the hypermedia topologies introduced by Parunak [18]. Simply speaking, subclasses of S-collection define a particular topology of encapsulated links, as can be seen in Figure 1. For the purposes of this paper, it suffices to describe the five predefined subclasses : Envelope, Folder, Menu, Freelinks, and Void:

1. **Envelope**: all members of an envelope are fully related, every member is linked to every other member.
2. **Folder**: an ordered set of members, each member having links to "next" and "previous" members.
3. **Menu**: a simple hierarchical structure; the head of a menu S-collection includes links to all other members, and each member is provided with a link to the head.
4. **Freelinks**: members of a freelinks S-collection may be arbitrarily connected by means of special INSERT_LINK and REMOVE_LINK operations.
5. **Void**: an S-collection devoid of internal structure, having only content (for example, "a" and "b" in Figure 1).

Classes 1 to 4 above are termed *complex* S-collections since they embody internal structure; class Void is termed *simple*. Complex S-collections may or may not have associated content; simple S-collections must have associated content.

Navigational Operations

Navigation within the HM Data Model is accomplished via four operations ACCESS, ZOOM_IN, ZOOM_UP and ZOOM_BACK, which are addressed to a particular S-collection.

All S-collections can respond to the message ACCESS. It implies executing the S-collection´s content (i.e., presenting some text, picture, audio, video clips etc.). Typically, a chunk of hypermedia information associated with the current S-collection is visualised on the user screen, but any kind of action can happen in response to the message ACCESS if the corresponding method has been overridden. Complex S-collections without content forward the message to their head.

Link following within the HM Data Model is simply a form of message passing. At any particular moment in time, the user can navigate only through a single, specific S-collection called the *current container*. Only members of the current container can receive messages during navigation. A concrete member of the current container is the *current member* for each particular navigational step. More precisely, the member that most recently received the message ACCESS is the current member. Only members related (linked) to the current member can be accessed (can receive the message ACCESS) in the next step of navigation.

Consider, for instance, navigation through S-collection "d" in Figure 1. S-collection "d" is the current container. If S-collection "a" is the current member, then it has been visualised (its content has been displayed by method ACCESS), and the user has links to members "b", and "c" available in the next step of navigation. In our prototype implementation, links emanating from the current member are depicted in the form of icons, buttons, clickable areas, etc. The user manually selects a link which results in the message ACCESS being sent to the corresponding member.

Since links are encapsulated within an S-collection, they become available for navigation only when the S-collection has been "entered" by means of the ZOOM_IN operation, which is available for all complex S-collections. The ZOOM_IN message is automatically addressed to the current member. For example, if the S-collection "c" becomes a current member during browsing current container "d" (see Figure 1), the user can apply the ZOOM_IN operation in order to make "c" the new current container. After a ZOOM_IN operation, the head of the new current container (i.e. S-collection "b" in this particular case) automatically becomes the current member and is visualised appropriately (i.e. receives message ACCESS).

Extending the functionality of ZOOM_IN to give access to any S-collection of which the current collection is a member is provided by the operation ZOOM_UP. Thus, if current member (say, S-collection "a") has been accessed during browsing the current container "d", then the user can "switch" to browsing of any other current container having the same S-collection "a" as a member (say, to browsing the container "c" in this particular case). Thus, the ZOOM_UP operation simply substitues the current container for another one leaving the current member unaffected.

The ZOOM_BACK operation is the complement of ZOOM_IN or ZOOM_UP. ZOOM_BACK restores the current container and current member to the state they had before the *most recent* ZOOM_IN or ZOOM_UP.

Together, ZOOM_IN, ZOOM_UP and ZOOM_BACK provide users with the capability of navigating in a direction orthogonal to the conventional plane of link-based browsing. Thus, we say that the HM Data Model supports an additional dimension of browsing hypermedia databases and extends the default browsing semantics.

Operations for Modifying the Database

New S-collection instances are created with the operation CREATE, whereby the name, content, and head are given as parameters. Once an S-collection has been created, new members can be inserted and existing members removed by means of the operations INSERT_MEMBER and REMOVE_MEMBER having a name of another S-collection (i.e. member) as a parameter. All links within S-collections belonging to the classes Menu, Envelope, and Folder are maintained automatically in accordance with their associated regular structure.

Of course, the regularly structured subclasses do not restrict us from using S-collections having arbitrarily connected members. Users can create an S-collection of class Freelinks and explicitly define its link structure using the messages INSERT_LINK and REMOVE_LINK whereby the source and destination are given as parameters.

The notion of links presented here should not be confused with that used in other hypermedia models, where links exist as independent fully-fledged persistent data objects. In our case, links and membership are simply mechanisms for reusing abstract data objects in different contexts via a special public interface. Notice that we do not map S-collections onto a global navigational level consisting of primitive nodes and links [1, 2, 21], rather we treat them as existing accessible chunks of hypermedia information.

All simple S-collections (those having content but devoid of internal structure) belong to class Void, where the operations ZOOM_IN, INSERT_MEMBER, and REMOVE_MEMBER are suppressed.

At any point, an S-collection can be deleted with the DELETE operation. In this case, its content and all links encapsulated within it cease to exist. If a deleted S-collection has been reused (defined as a member) by other S-collections it is removed from S-collections by means of the REMOVE_MEMBER operation (see Figure 1).

Thus, the traditional linking of chunks of multimedia data used by nearly all contemporary hypermedia data models, is replaced with the insertion/removal of S-collections into/from the private memory of other S-collections. Note also that in the HM Data Model, membership is not restricted to a hierarchical structure, recursive membership is both possible and meaningful.

Illustrative Example

As an example of the application of the HM Data Model, consider the situation in Figure 2. Suppose that we would like to develop a hypermedia presentation of our institute, the IICM. A typical hypermedia database according to the HM Data Model might represent the institute as a whole as the S-collection "IICM" in Figure 2.

The content (data unit) of the S-collection "IICM" might typically include the institute logo and a short textual description of the collection (e.g. "Welcome to IICM. Click now on Zoom In in order to get started"). Suppose that "IICM" is an S-collection of type Menu and has other members "Projects", "Personnel", etc.

If "IICM" is the current container and the member "Personnel" which is an S-collection of type Folder, is accessed (clicked), its content (e.g. the text "This collection consists of personal files of IICM employees. To look through them one-by-one click now on Zoom In") is visualised. If the user now clicks on button "Zoom In", the first file (i.e. the content of S-collection "Hermann") is shown; the remaining files can be traversed by clicking buttons "next" or "previous" (the navigational aids available in a folder). The operation ZOOM_UP provides access to any S-collections of which the person is a member (say, to the S-collection "Hyper-G" in this particular case). Clicking on button "Zoom Back" returns the user to S-collection "IICM" and its navigational paradigm. Now clicking on another member of "IICM" (say, on "Projects") visualises that member's content, clicking on button "Zoom Back" leaves "IICM", and so forth.

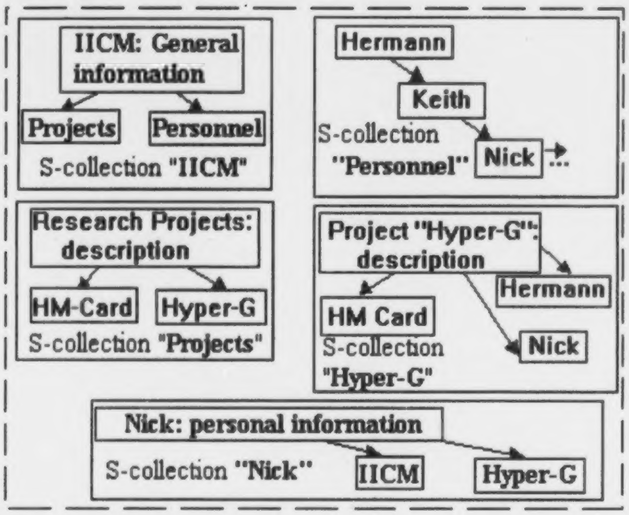


Figure 2: Example of a Database Structured Using S-collections

This situation is typical of the navigational process in the HM Data Model and should be clearly understood: when a user accesses an S-collection *B* as a member of S-collection *A*, the navigational paradigm of *A* continues to be available. Only when *B* is entered (operation ZOOM_IN), is the navigational paradigm of *B* activated, and the navigational paradigm of *A* becomes unavailable. When now selecting a member *C* of *B*, *B* and *C* start to play the roles of *A* and *B*, respectively. Of course, the ZOOM_BACK operation allows backtracking to the navigational paradigm of *A*.

A hypermedia description of a particular person (say "Nick") can be enhanced with S-collections describing projects this particular person is involved with and an S-collection describing his/her affiliation (say, "Hyper-G" and "IICM" in this particular case). Further, information about employees engaged in a particular project can be inserted into a corresponding S-collection as members (say, "Hermann" and "Nick" in this particular case). Of course, all S-collections describing different projects can be combined into an S-collection "Projects".

The property of the HM Data Model that an S-collection may belong to many other S-collections, even recursively, provides all the necessary power to deal with more complex situations. Thus, for instance, if it is desirable to have references to descriptions of all projects sharing particular equipment within the description of this equipment, the corresponding "Project" S-collections are simply inserted into the S-collection "Equipment". Note the recursive membership of S-collections "Hyper-G" and "Nick" in this example. Such recursive membership elegantly handles the common situation where a user needs to access information about "Hyper-G" while browsing all information related to "Nick", or vice versa. Moreover, if a certain project (say, "Hyper-G") is based on other projects (say, "HM-Card"), then the description of the S-collection "Hyper-G" can be extended with the relevant members (with the S-collection "HM-Card" in this particular case).

Note that the concept of encapsulation of links makes this process extremely flexible. Any S-collection can be inserted into any other S-collection (message INSERT_MEMBER) without considering the links directed to or emanating from it: new links having well-defined context (that of the containing S-collection) are added automatically; additional links can be added manually if desired (method INSERT_LINK which can be inherited from the system class Freelinks). The asymmetry of most hypermedia models (some objects belong to one part of the database and other parts just link to it) disappears: an abstract object belongs in exactly the same way to all containers, wherever it is relevant.

Algebraic Operators

Algebraic operators are applied to S-collections and produce a new resultant S-collection which becomes the current container. Thus, these operators provide users with the possibility to adjust the current scope of link-based navigation to their particular *short-term* needs. The algebraic operators broaden the concept of current container (see Section 2.2) and serve as an additional tool for browsing hypermedia composites.

The operator INCLUDE is applied to the current container and one of its members; the operator temporarily replaces the member with its internal navigable structure. Thus, for instance, if the user INCLUDEs the members "Projects" and "Personnel" during browsing current container "IICM" (see Figure 2), the scope of link-based browsing is extended; and the resultant current container is shown on Figure 3a.

The following simple rules are used in order to avoid possible ambiguity:

1. Links emanating from (or directed to) a member which has been replaced with its internal structure, emanate from (or lead to) its head (see Figure 3a);
2. It is possible for an S-collection to be both a member of the current container and the included S-collection. In this case, the S-collection appears only once (is not duplicated) with both sets of links. For instance, a current container which is a result of INCLUDing the S-collection "Hyper-G" in the current container "IICM" (see Figure 3a) is shown on Figure 3b.

Thus, the current container has been composed from the navigable structures of a set of S-collections. Such set of S-collections is called a *perspective*. For instance, the perspective for the current container shown on Figure 3b, consists of the S-collections "IICM", "Personnel", "Projects" and "Hyper-G".

The operator JOIN is applied to a current container and another S-collection sharing a common member, the operator simply extends the current container with the internal navigable structure of the second S-collection. Thus, for instance, if the user JOINs the current container "Projects" (see Figure 2) and the S-collection "Nick" sharing a common member "Hyper-G", the scope of link-based browsing is extended; and the user continues navigation in the current container shown on Figure 4.

The operator DIVIDE can be applied to the current container and one of S-collections belonging to the perspective. The DIVIDE operator is the complement of the Include (or JOIN) operator which was applied to the selected S-collection in order to include it into the perspective. Thus, for instance, if the user DIVIDEs the current container shown in Figure 3b by the S-collection "Hyper-G", the scope of link-based browsing is reduced and navigation continues in the container shown in Figure 3a.

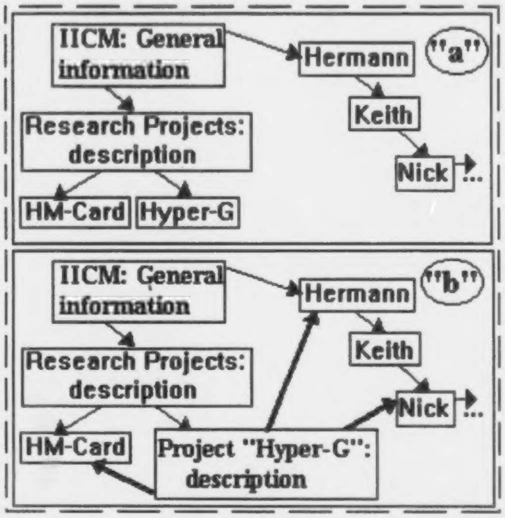

IICM: General information
Hermann
"a"
Research Projects: description
Keith
Nick
...
HM-Card
Hyper-G
IICM: General information
Hermann
"b"
Research Projects: description
Keith
Nick
...
HM-Card
Project "Hyper-G": description

Figure 3. Current container extended by the Include Operator

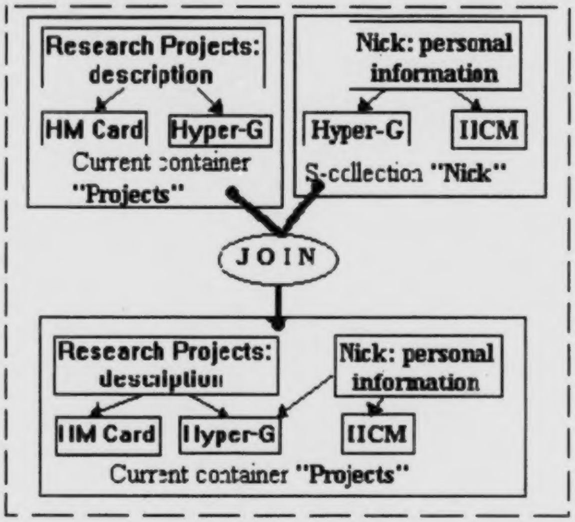

Research Projects: description
HM Card
Hyper-G
Current container "Projects"
Nick: personal information
Hyper-G
IICM
S-collection "Nick"
JOIN
Research Projects: description
Nick: personal information
HM Card
Hyper-G
IICM
Current container "Projects"

Figure 4: Current Container Extended by the JOIN Operator

In our prototype implementation, all the algebraic operators are available in the form of special buttons. Pressing a particular button invokes a list of S-collections which can be parameters of the operator in accordance with its semantics and current browsing situation. The user manually selects an S-collection which results in the modifications to the current container. Overview diagrams coupled with a display of the current perspective provide all the necessary feedback about the user's current location.

Concluding Remarks

The HM Data Model is a hypermedia data model based on abstract data objects called S-collections, which represent reusable "chunks" of hypermedia information (of any size). An S-collection may embody internal structure: hypermedia links are encapsulated within an S-collection and are bound between its member S-collections. The HM Data Model defines not only data structures (i.e. S-collections) per se, but also defines operations applicable to instances of such data structures and extends the default browsing semantics by introducing a new dimension of navigation.

The HM Data Model is distinct from other hypermedia data models in the level of granularity of the hypermedia data. Thus, S-collections are fully-fledged linkable and *navigable* objects. An S-collection encapsulates a particular link-based browsing strategy which is not mixed with strategies encapsulated within other S-collections (until users explicitly combine them by means of the algebraic operators). S-collections are well-defined units of interaction. A particular S-collection can be created, modified, deleted and/or reused as an independent abstract object via its public interface, without taking into account the internal structure of other S-collections or the hypermedia database as such. While a similar level of granularity is also provided by the Aquanet model, the authoring and browsing mechanisms described in this paper are unique to the HM Data Model.

The HM Data Model has been implemented as a prototype system for MS-Windows called HM-Card. The software, its documentation, and the implementation of the example described in this paper are available via anonymous ftp from "iicm.tu-graz.ac.at" in directory "pub\hmcard15".

Internet India

Suzanne McMahon
University of California Berkeley Libraries U.S.A.
smcmahon@library.berkeley.edu

Abstract: India has had Internet access since 1988 when ERNET users were connected live to the Internet through a link to UUnet in the USA. In August 1995 VSNL (Videsh Sanchar Nagar Ltd.), currently the sole providers for international communications in India, offered both text and graphic connectivity to the Internet. Penetration of phone lines is still a limiting factor, but Internet use in India is growing. The article lists current web sites created in India or sites on Indian topics.

The Internet in India

On a recent trip to India I attended the New Delhi World Book Fair and was struck that at that Fair two paradigm shifts for information delivery were going on simultaneously. Clearly the Book Fair was promoting the shift from illiteracy to literacy or in this context it might be more appropriate to say the shift from oral tradition to print culture. In a country where 48% of the population is still illiterate, according to the 1991 Census of India, the magic of the book and the power that reading imparts still has tremendous resonance. Simultaneously, the Book Fair was hosting a workshop on digital libraries in the Internet age and a teleconference on "Doing Business on the Internet," heralding the shift from print tradition to digital culture and the passing the magic wand on to a new media.

In a global environment the unique information-age challenges of the developing world are the challenges of the world community. Al Gore's vision of the GII (Global Information Infrastructure) of fiber optic-based telecommunications that will bring about economic growth, strong democratic participation, environmental responsibility, and shared stewardship for the fragile economic, political, and environmental systems of the world community may be quite a stretch for countries like India, at least in the near future. While it is feasible in the U.S., home to half the Internet users of the world, to talk about curbside delivery of fiber optic-cable, to most of the developing world achieving a penetration rate of phone lines above 5% presents a challenge. In a world where information is arguably becoming the single most important commodity the danger of splitting the world into information-rich and the information-impoverished is very real. The ability to stay aware of the "thin" end of the net while planning for global information delivery is an aspect of social conscience that policy makers and planners must develop. The November 1995 issue of Internet World focused on the Net in Europe, Africa, and China. The following paper focuses on the Net in India. India has had Internet access since 1988 when ERNET users were connected live to the Internet through a link to UUnet in the USA. ERNET had been established by a UNDP (United Nations Development Projects) grant to link premier educational and research institutions in India: The IITs (Indian Institutes of Technology), the IIMs (Indian Institutes of Management), NCST (The National Center for Software Technology) and the Indian Institute of Science (IISc) were linked together in 1988, and, through a link to UUnet in the USA, were connected live onto the Internet. But participation by a broader audience has been limited, until very recently, by cultural, political and economic factors within the country.

Restrictions from the Indian Department of Telecommunications effectively kept most players out of the Internet game. The traditional role of DoT in socialist India had been regulating rather than enabling and the government viewed free flow of information as a security threat rather than as a necessary precondition for political, social, and economic health. The dismantling of the Soviet behemoth triggered a fundamental reorientation for India, as well as for many other former Soviet allies. Economic liberalization in 1991 signaled significant changes in government policy and spurred unprecedented economic growth in the country. On the information front a number of Indian and foreign vendors began to offer e-mail accounts. In 1994 basic telecom and value-added services in India were privatized and India began an ambitious program to increase the number of telephone lines and other means of communications throughout the subcontinent. In August 1995 VSNL (Videsh Sanchar Nagar Ltd.), currently the sole providers for international communications in India, offered both text and graphic connectivity to the Internet. Even in the privatized environment, VSNL is a virtual monopoly with extensive concession from DoT. However, recent telecom legislation, Supreme Court rulings, and pressure from the business community will almost certainly alter VSNL's privileged position.

Access to electronic information over the Internet in India is limited by the degree of penetration of phone lines in the country, a little over 1% (around 10 million lines), with a goal of expanding the penetration to 3% by the year 2000. Demand far outstrips supply. The waiting list for new connections in 1993 was 2.8 million and waiting periods of months, even years, have been the norm. India's goal is to provide phone installation on demand by 1997. However, India's attempts to attract foreign investors to create a needed upgrade in the nation's telecommunications infrastructure have been stymied, to some degree, by the country's licensing regulations. The installed base of computers is another limiting factor for India. Currently India has about 650,000 computers, less than the installed computer base in San Francisco. Social resistance to computerization by business and labor groups, fueled by fear of losing control of traditional markets or fear of losing jobs has been a heavy drag on efforts to computerize and modernize operations, the banking industry being a striking example. But attitudes are changing rapidly.

One hopeful sign is that India is concentrating on leap-frogging out-dated technology steps as it establishes its new infrastructure, concentrating on fiber and satellite connections. The launching of India's INSAT2C telecommunications satellite is a major step forward for Internet infrastructure. It is fitted with three Ku-band transponders for fixed satellite service, which covers all major cities of India. The signal is so powerful that it can be received by a 1.5 ft diameter rooftop dish, which opens a whole range of possible uses by the public and private sector. An S-band will make possible wide-range transmission of radio and television signals, not only in India but also in the Middle-East and South-East Asia. Besides the three Ku-bands, two C-Band channels provide extended coverage. This means live Internet connection for major cities and store and forward email for more remote rural areas. I told one software developer about the new telecommunications satellite and email for the villages and he replied "Email? The villages don't need email; the villages need water." Of course, this is true, but in a global society where information is becoming the most important commodity, the boost to rural development through advanced information delivery systems is incalculable. .

Web Sites about India General Sites

Web pages on India that include general, comprehensive lists linking to other web pages offer an excellent overview of what's available. The most comprehensive list of web resources for India is the World Wide Web Virtual Library: India (http://webhead.com/WWWVL/India/) maintained by the Australian National University. The page, which is updated frequently, is divided into subject categories including: Art and culture, History, Internet in India, Language, Literature, Maps, Music, organizations, People, religion, Sports, and States. A periodic check of WWWVL:India will keep you up-to-date on most of what's new and interesting on the web for India. Yahoo (http://akebono.stanford.edu/yahoo/Regional_Information/Countries/India/) also provides links to a wide variety of resources on India.. Search tools like Alta Vista (http://altavista.digital.com/), Inktomi (http://inktomi.berkeley.edu/), and Excite (http://www.excite.com/) are constantly being improved and turn up an abundance of material through keyword searching.

Online periodicals

Electronic publishing has begun to blur traditional print-based definitions of format and it is not easy to describe the qualities of an online periodical. In this section I have included web pages that identify themselves as electronic periodicals. A few Indian periodicals and newspapers that publish in print also publish an electronic version. The Hindu, a leading national newspaper, puts up its weekly edition (http://www.webpage.com:80/hindu/), as well as an edition of Business Line, a weekly business news magazine (http://www.indiaserver.com/news/bline/bline.html), up on the web. Electronic Darpan (http://www.sponsor.net/~pmishra/darpan/) is the online version of Darpan, a literary magazine, published by students at the University of Illinois at Urbana-Champaign, that explores the rich Indian and Indian-American experience with featured prose, poetry, and graphics. Publishers from the

Indian-American ethnic press are also putting up web versions of their magazines. for instance, India Currents (http://www.rahul.net/indiacur/). Some periodicals are offered only online. IndiaWorld (http://www.indiaworld.com/), a general magazine published in Bombay, includes shopping opportunities, job listings, matrimonials, and business news. IndoLink (http://www.indolink.com/), published in San Jose, California, has calendars of local events, general and economic news, recipes, and a humor column. CyberINDIA (http://www.cyberindia.net/cyberindia/main/indiacen.htm) offers articles on business, sports, literature, food, and travel. Several scholarly journals are issued by universities only in electronic form. Sagar (http://asnic.utexas.edu/asnic/sagar/sagar.main.html) is a biannual journal sponsored by the Center for Asian Studies at the University of Texas at Austin. The journal showcases exceptional graduate student and junior scholar articles . The Journal of South Asian Women's Studies (http://www.shore.net/~india/jsaws/) , currently located at Harvard, publishes theoretical and practical papers on issues that are of interest both to scholars of South Asia and to women in and from South Asia. Specific subjects include gender issues, religion, philosophy, politics, the arts, discoveries, and cultural or social products by women.

Language and Literature

Language resources are still limited on the web. The Human Languages page (http://www.willamette.edu/~tjones/Language-Page.html) gives a good overview of what's available. For instance, the University of Pennsylvania Web site for Hindi (http://philae.sas.upenn.edu/Hindi/hindi.html) offers simple audio lessons. A short description of most Indic languages can be found at Ethnologue (http://www-ala.doc.ic.ac.uk/~rap/Ethnologue/eth.cgi/India/). The WWWVL:India lists links to assorted language pages (http://webhead.com/WWWVL/India/india210.html). For classical Indic languages the best source of information is the Indology page (http://www.ucl.ac.uk/~ucgadkw/indology.html). Indology and the South Asia Gopher Archives (gopher://gopher.cc.columbia.edu:71/11/clioplus/scholarly/SouthAsia/Texts/Archives) both include lists of electronic text archives.

Art and culture

A good example of the high quality art exhibits that are showing up on the web is the Arts of India Multicultural Art Print Series (http://www.artednet.getty.edu/ArtsEdNet/Tools/136/136_2.html). Many state and city pages provide good mini-tours of parts of India, for instance the Rajasthan page (http://www.ent.ohiou.edu/~kartik/raj.html). The Indian Humor Page (http://www.rajiv.org/ii/) includes humorous dialogue from Hindi films, regional jokes, and humorous essays. The Indian Last Names page (http://www.nssl.uoknor.edu/~lakshman/names.html) describes how Indian last names are derived--from caste, town, religion, and so forth--and why the last name functions differently in India than in the West. The South Asia Diaspora Web page (http://www.lib.berkeley.edu/SSEAL/SouthAsia/diaspora.html), includes essays on diaspora topics, bibliographies, photographs, and links to other electronic resources on the South Asian Diaspora.

Business sites

The Indian business community, sometimes in collaboration with foreign investors, has put up web pages with information and news about doing business in India. One good page with a U.S. orientation that draws together quality sites on the subject of business in India is Indiaserver (http://www.indiaserver.com/ilink.html). The India Biggest Emerging Markets (BEMS) page is maintained by U.S. Department of Commerce, Economics and Statistics Administration, and gives information on India for U.S. exporters (http://www.stat-usa.gov/bems/bemsind/bemsind.html). JITNet, The Japan-India Technology Network (http://sunsite.sut.ac.jp/asia/india/jitnet/index.htm) was created to serve as a communication link between technologists, researchers, academicians, businessmen, and other individuals on issues related to science and technology in Japan and India. The site includes overviews of Indian economy and business practices; technology updates; links to Software Technology Park web pages and newsletters; World Bank projects in India; and a bibliography of articles on economic reforms.

World Wide Web Resources on India Servers

The National Informatics Center has put up a web page (http://www.nic.in/) describing its program of state-of-art solutions to the information management and decision support requirements of the Government and the Corporate sector, including conducting feasibility studies for computerization and designing, developing, and implementing computer-based information systems.

Besides information about the ERNET program, the ERNET page offers news from India, updates on Indian cyberspace, and links to Indian web pages, for instance, the Unesco-India Home Page (http://www.doe,ernet.in/~unicefd/) describing the Indian Unicef program. Universities and research institutions that have access to ERNET have built home pages describing their programs and resources.. One good example is the IISc in Bangalore (http://ece.iisc.ernet.in/iisc.html). View from the Ground (http://ece.iisc.ernet.in/monster/atul/indianet/) offers an irreverent description of the development of the Internet in India, links to current articles about the Indian computing scene, and free software for Indian Internet users. Indiagate (http://www.indiagate.com/bbs/bbslist.html) gives a complete list of Indian BBSs with instructions on how to sign up, and an article describing the "typical" Indian BBS user.

Gopher

Gopher technology has been around longer than the web and some universities have developed very extensive gopher sites. For India one of the oldest, best-known, and most useful is the South Asia Gopher (gopher://gopher.cc.columbia.edu:71/11/clioplus/scholarly/SouthAsia/). The South Asia Gopher is collection of world-wide network accessible information resources relating to South Asia. It includes direct links to online catalogs of the world's top library collections on South Asia; lists of South Asia related newsgroups, listservs, BBSs, organizations, societies and other NGOs; and the International directory of South Asia Scholars. A web version of the South Asia Gopher is being developed.

Newsgroups and Listservs

Currently there are over 50 newsgroups and hundreds of listservs dealing with subjects related to India. Through newsgroups and listservs it is possible to participate in discussions of current interest topics and to stay updated on new Internet resources. Newsgroup FAQs provide answers to frequently asked questions and the listserv archives of previous discussions can also be highly useful resources in themselves. A comprehensive list of newsgroups and a list of some of the better known listservs is available at the WWWVL:India page (http://webhead.com/WWWVL/India/india5.html). The short list below illustrates some of the major categories.

Newsgroups

NAME	SUBJECT
alt.culture.us.asian-indian	Asian Indians in the US and Canada
alt.movies.indian	Films from India and the Indian subcontinent
misc.news.southasia	News from Bangladesh, India, Nepal, etc. (Moderated)
rec.music.indian.misc	Indian music in general
soc.culture.indian	India & things Indian
soc.culture.tamil	Tamil language, history and culture
soc.religion.islam	Islamic faith (Moderated)
soc.religion.sikhism	Sikh Religion and Sikhs all over the world (Moderated)

Listservs

INDIA-IN-LANGUAGES@EE.ROCHESTER.EDU
Scholars & speakers of Indian languages Subscription requests to INDIA-INLANGUAGES-REQUEST, not LISTSERV.

INDIA-D@INDNET.BGSU.EDU
Daily discussion of news digests in English

INDIA-E@INDNET.BGSU.EDU
Economic Forum at BGSU

INDIA@CUNYVM
News and discussion, some emphasis on computers, includes "The India Journal" and "South Asia Currents"

INDIA-ALERT@CS.WISC.EDU
News about human rights in India

INDOLOGY@liverpool.ac.uk
Forum for Sanskritists, MIA and NIA language specialists, Dravidologists, historians, and others interested in any aspect of Indological studies

TAMIL-L@DHDURZ1
Tamil Studies

Conclusion

This article has reviewed the current state of the Internet in India and has presented a small sample of Internet resources available on the World Wide Web and through listservs, and newsgroups. It is possible to locate other resources and stay updated on new resources by periodically checking web pages that maintain comprehensive lists of links to resources about India, doing keyword searches on the ever-improving search tools available, and by following announcements and discussions on listservs and newsgroups.

References

Journal References

[Adonis 1995] Adonis, Andrew. (1995). Oxcarts on the Infobahn, World Press Review v42, n1 (Jan, 1995):40.

[1995] Clearing the line: Indian telecoms, Economist v337, n7940 (Nov 11, 1995):62 (2 pages).

Interviews

A large part of my article was based on my recent trip to India and interviews I conducted while in New Delhi, the Punjab, Madras, and Bangalore including: Tejeshwar Singh, Sage Publishing; Dr. Mvvs Reddi, National Informatics Center; Dr. H.K. Kaul, Delnet; Dr. A. Lahiri, NISSAT; Dr. T. Viswanthan, INSDOC; Dr. Harish Chandra, IIT, Madras; Dr. N.M. Malwad, IISc, Bangalore.

Virtual Office Hours: A Communication Tool for Students and Teachers

Professor Craig A. Merlic*
Department of Chemistry and Biochemistry
University of California, Los Angeles
Los Angeles, CA 90095 USA
merlic@chem.ucla.edu

Matthew J. Walker'
Department of Chemistry and Biochemistry
University of California, Los Angeles
Los Angeles, CA 90095 USA
mjw@chem.ucla.edu

Abstract: The design and implementation of a World Wide Web (WWW) based instructional tool titled Virtual Office Hours (VOH) is described as implemented in the Department of Chemistry and Biochemistry at UCLA. The project is designed to facilitate communication between faculty and students through on-line access to all instructional materials and on-line question and answer capabilities. The project serves more than fifteen classes and three thousand students per term and it has generated more than 80,000 access hits in a single week of usage.

Introduction

The tremendous amount of public and commercial attention recently focused on the Internet [Segal 1995][Zakon 1996] and the World Wide Web (WWW) [Berners-Lee et al 1992] has stimulated interest in applying these technologies to improving educational services for students at academic institutions. At UCLA, this process has begun in earnest in the Department of Chemistry and Biochemistry. The WWW provides an excellent resource for facilitating student-faculty communication and for demonstrating graphical concepts in ways that are impossible with simple paper handouts or chalkboard demonstrations. The WWW provides today the ability to use full color real-time animation, extensive graphics, and hyperlinks to other resources. In the future, resources like Virtual Reality Modeling Language (VRML) [Bell & Parisi & Pesce 1996] will allow for full 3D representation of images moving in space. An important aspect of Internet delivery of instructional materials is that they can be shared worldwide, thus resource rich schools can assist less endowed schools. The goal is to enhance the learning potential for all students.

Project Design and Implementation

The Virtual Office Hours (VOH) project was conceived as a means for providing two principal lines of communication between faculty and students. First, instructional materials such as announcements, syllabi, practice old exams, exam keys, problem sets and handouts are placed on-line, organized by class, for remote access by students. This solves problems with lost or missed handouts, limited library hours, and checked out reserve materials. Second, students and faculty can directly communicate in an open forum, much like that of a regular office hour, but with many advantages. In particular, all questions and answers are publicly posted for the benefit of all students. Therefore, unlike traditional office hours where only those students who attend benefit from the student-faculty discussion, on-line forums can benefit all students in the class by making a written record of the dialog available on a twenty-four hour basis. For the instructor, this method provides relief from having to answer the same or similar questions repeatedly in several office hour sessions. For students, access to questions on-line allows for better preparation for lecture and regular office hours, elimination of time conflicts, and the ability to post questions as they arise during studying.

Instructional Materials

The first component of VOH presents large amounts of textual and graphical data in an organized, usable, and easily accessible manner. Important issues include organizing the documents in such a way as to make access intuitive and quick, allowing for easy maintenance, accepting all types of documents from faculty, and keeping documents as true to the original as possible while minimizing data size.

A student arriving at the main VOH home page, finds three sections: the general information section, the class pages section, and the outside links section. The general information section includes links to an archive of exam pages which allows previously submitted exams to be searched and displayed; an instructions page which provides links to instructional documents specific to VOH and to on-line use guides for email, the Internet, etc.; a comments form; VOH from previous terms; a distribution section [Distribution Area], where all of the scripts that run the project can be found; and the "white page", or description of the project. The class section divides the term's classes into their respective fields of chemistry: General, Physical, Organic, Biochemistry, and Inorganic. These fields are displayed in a tabular format along with the specific classes available therein. Clicking on a field header takes the student to a page with links of interest to that specific field of chemistry. The final section, for outside links, provides access to information about the Department of Chemistry and Biochemistry, the main UCLA home page, and several other educational projects on the WWW that are relevant to chemistry.

A student pursuing the link to a specific class from the VOH home page is presented with the "basic class package" of documents. This includes: a link to departmental information about the instructor; a syllabus for the class; a frequently asked questions (FAQ) section comprised of the edited consolidated questions and answer sections from previous terms of the class; an announcements area [Submitting Announcements]; and a questions and answer (Q&A) section, see [Questions and Answer Section]. In addition, instructors may submit information such as old examinations, examination keys, handouts, problem sets, supplemental readings, animations, and links to outside resources, all of which can be added to their page to improve its usefulness to the student.

Because of the large number of files necessary for each class, a rationally designed directory structure is crucial to the facile maintenance of the project. VOH's is simple and consists of a "root" directory called *uclavoh* with the following subdirectories: *docs*, for documents; *profs*, a secure area for instructor only access; *conf*, another secure area for configuration materials; *images*, for images on all project pages such as backgrounds and navigation icons; and *class*, where *class* is the designation of a class in the department supported by the VOH project. Each class directory contains subdirectories for images and questions along with its home page. Because of the consistent and modular nature of the directory structure, all of the class directories can, at the end of the term, be moved into a different directory, such as *winter96*, and maintained as a whole functional piece while new class directories are created under the project root for the current term. Hence, the current classes are always found directly under the root, *uclavoh*. To keep the Universal Resource Locator (URL) for the class as short and easy to remember as possible, we take advantage of a feature of the NCSA httpd server (version 1.4.1) [SDG 1995]. If a directory is specified in a URL, the NCSA server's index function will do one of two things: if a file called "index.html" is present, it will be displayed, otherwise an "ls-like" index of the directory will be generated and displayed. Therefore, if we name the VOH main home page "index.html" and place it in the root directory, the project can be referred to by the concise URL: http://www.chem.ucla.edu/uclavoh. The same holds true for individual classes which can be referred to by their directory name, e.g. http://www.chem.ucla.edu/uclavoh/10A.

The greatest difficulty in setting up class pages for instructors is reconciling the various formats of their existing data with the formats necessary for the WWW. Old midterm examinations are a particular challenge because of the wide variety of data formats (typewritten, hardcopy with "cut and paste" graphics, handwritten, or digital in one of many possible formats) and the frequency of special characters, special math symbols and graphical images, which are not directly supported by the WWW browser for viewing. We use a "text if possible, graphical interchange format (GIF) [CompuServe 1989] otherwise" policy. If the information to be presented is primarily textual, such as a Microsoft Word [Microsoft 1996] document or a text file, it is manually converted directly to hypertext markup language (HTML) [Connolly 1995]. Otherwise, the document is converted into GIF graphics on a page by page basis. This was accomplished using two pieces of software. First Print2PICT [Raoult 1993] was used to capture print output as PICT format graphics. These were then converted with GraphicConverter [Lemke 1995] to GIF graphics. Old materials and handwritten materials were scanned with an Apple ColorOne scanner (at a resolution appropriate for viewing on the screen) and converted to GIF graphics. These GIF graphics are then individually "encapsulated" by automatically written HTML documents to provide for easy printing and navigation.

Question and Answer Section

Electronic student-instructor dialog can be accomplished simply by advertising the instructor's email address, but this method has several flaws. First, students as a group cannot benefit from a discussion that is conducted via personal email; second, the professor would be obliged to answer redundant questions inadvertently asked by multiple students; and third, this could produce an undue volume of additional email for the instructor. Therefore we created the second major component of the project, a form-based question and answer forum. Students clicking on the "submit question" link of their class's Q&A section are presented with an HTML fill out questions submission form (qsf). This form prompts them for their name and email address, and provides them with a field for entering a text question. The student may submit questions either for public consideration or "confidential" posting to the instructor. A public post is immediately accessible to all students through the "unanswered questions" link from the Q&A section of their class home page, thereby increasing the usefulness of asked questions, and reducing redundancy. The confidential channel still allows for private questions to be asked of the professor. Answered questions are viewed by selecting the "answered questions" link in the Q&A section.

To implement this strategy we designed a Perl [Wall 1991] common gateway interface (CGI) [CGI] script. The question submission form (qsf) script first verifies that the form has been properly filled out, and formats the question for output. Public questions are each saved as a separate file in the class's questions directory, named as the UNIX time (the number of seconds since January 1, 1980 [Wall & Schwartz 1992]) at submission. This nomenclature allows the questions to be concatenated in chronological order for viewing by another Perl CGI script (questions to be answered, qtba). Confidential questions are emailed directly to the instructor using the UNIX mail program. A serious security issue arises when passing form data directly from an HTML fill out form to a UNIX program. Because certain UNIX programs (such as mail) can send commands directly to the system, a student could inadvertently send a dangerous command to the system (such as the command to remove all of the files on the hard drive!) To prevent such a disaster from occurring, the form data is parsed and dangerous characters are removed.

Answering Questions

An important element of making the VOH project viable in the Department of Chemistry and Biochemistry at UCLA is making the project very easy and convenient to use by instructors. Our goal is to improve their ability to instruct without imposing significant additional demands on their time. To accomplish this goal, we created a script that allows the instructors to edit and answer student questions and post announcements for their class, without any knowledge of HTML. The instructor only needs to request a document (the questions answering and announcement posting form or "qaarf") in the *profs* directory, provide their username and password, and fill out this simple form. The form then asks to which class to post the information. Upon selecting "answer", the script retrieves the oldest unanswered question from the proper class's question directory and presents it in an editable field of a new form. Here the instructor may answer the question by simply typing in a reply into the answer field, or he/she may skip or delete the question. Upon submission of an answer, the script creates the necessary HTML and prepends it to the list of answered questions in the class's questions answered (qa) file. This prepending maintains a reverse chronological (most recent first) order of the answered questions, allowing newly answered questions to be quickly located by the student at the beginning of the qa file.

Submitting Announcements

One of the more difficult things for an instructor to do is circulate an announcement to the class in a timely manner, making sure that even absent students are notified. Because posting an answer to a question and posting an announcement are logically very similar, we modified the qaarf form to include an announcements option. Upon selecting "announcement" from the qaarf form, a new form is generated which prompts the instructor for the announcement text. Again the HTML is automatically generated and the prepended the class's announcements file, which is directly accessible by the students from the class page.

Reminder Robot

For the Q&A section of Virtual Office Hours to be successful, questions submitted by students must be answered in a timely fashion. However, because there is generally no need for the instructors to look at their class pages, it is inconvenient for most instructors to check for unanswered questions on a daily basis. Therefore, we developed a simple Perl script which, when executed on a daily basis by the UNIX cron utility (a program that executes commands at specified times), checks for unanswered questions and reports the number and age of waiting questions as an email message to the individual faculty. We find this vastly decreases the amount of time that instructors must spend using the system, and decreases the time between posting of questions and answers. The instructors appreciate only needing to access the system when notified of waiting questions, and students find a majority of their questions are answered within twenty-four hours of being posted.

Results

Quantifying the impact of a WWW site is a topic of much interest today and very soon we can expect to see better, more accurate tracking software emerge. Currently, however, we have applied only simple methods to examine the impact of our service at UCLA.

The first method is to simply consider the number of hits on our server due to access to the project. The Department of Chemistry and Biochemistry WWW server went into full operation in October of 1993 and receives approximately 20,000 hits per week. The VOH project began full operation in September of 1995 and resulted in a dramatic increase in server hits, rising to more than 100,000 in some weeks [Fig. 1][Wusage 1994]. It is interesting, but not surprising, to note that there are peak usage periods before midterm and final examinations as students access course materials. One could envision using VOH access patterns to study student study habits.

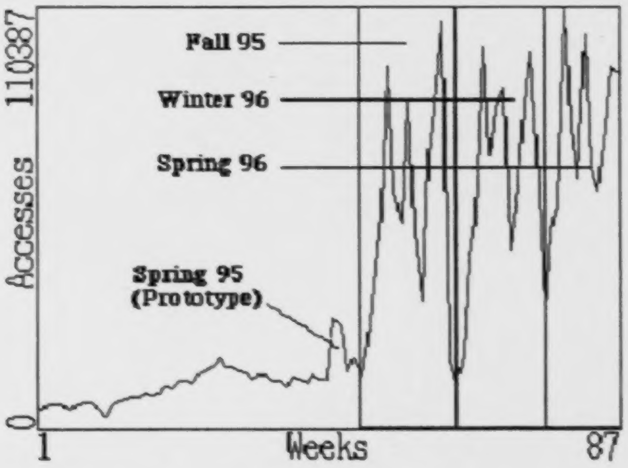


Figure 1: Accesses to the UCLA Chemistry and Biochemistry WWW server.

The second method that we used to measure our impact was to conduct a voluntary user at the end of school terms. The survey asks students questions about their usage of the VOH program and its usefulness for their studies. Many students responded that their typical usage was once or more per week and that they found it very useful as a supplemental educational resource. Sample student comments include:

"I think this is great. It's a wonderful resource for chemistry students."

"I get invaluable class material from VOH and I get to talk with the professor despite my busy schedule that doesn't allow for regular office hours attendance."

"This is a beneficial augment to live office hours in that each of us is able to access the questions and answers of all other students in the class."

"I sincerely hope that VOH will eventually be able to incorporate all courses and professors on the Web."

The completed project can be viewed at http://www.chem.ucla.edu/uclavoh. It is open for worldwide viewing.

Distribution Area

A link is provided on the main VOH home page to our "distribution area", which allows other educators to download our project materials. The main scripts that run the project can be found here along with examples for HTML documents, maintenance scripts and instructions for setting up a VOH site. We are currently working on porting these scripts to the Macintosh and Windows NT and these versions of the scripts, along with instructions, will also be available in the distribution area. See: http://www.chem.ucla.edu/uclavoh/dist.

Conclusions

The World Wide Web is an outstanding tool for the presentation of instructional materials to large numbers of students and for enhancing student-faculty communication. It crosses boundaries of all types by allowing students from all over the world benefit from work done by a small number of people. Posted class information not only benefits UCLA students, but can also be used as study aids by college, university, and even high school students worldwide. Student-faculty communication is promoted with many new advantages over traditional office hours.

References

[Bell & Parisi & Pesce 1996] Bell, M.; Parisi, A.; & Pesce, M. (1996). The Virtual Reality Modeling Language. Version 1.0C, http://vag.vrml.org/vrml10c.html.

[Berners-Lee et al 1992] Berners-Lee, T. ; Cailliau, R.; Groff, J.F.; Pollermann, B. (1992), "World-Wide Web: The

Information Universe", *Electronic Networking: Research, Applications and Policy*, (2) 52-58. Spring 1992, Meckler Publishing, Westport, CT, USA.

[CGI] Common Gateway Interface. http://hoohoo.ncsa.uiuc.edu/cgi/interface.html.

[Connolly 1995] Connolly, D.W., (1995). Document Type Definition for the HyperText Markup Language. http://www.w3.org/pub/WWW/MarkUp/html-spec/html.dtd.

[CompuServe 1989] CompuServe (1989). Graphic Interchange Format.

[Lemke 1995] Lemke, T. (1995). GraphicConverter. Version 2.3. Lemke Software, Erch-Henkel-Ring

8a, D-31228 Peine, Germany.

[Microsoft 1996] Microsoft Co. (1996). A majority of our users use some version of Microsoft Word for the Macintosh, such as: Microsoft Word Version 6.0.1 for the Power Macintosh, 1996. http://www.microsoft.com/msword/.

[Raoult 1993] Raoult, B. (1993). Print2PICT. Version 3.5.

[SDG 1995] Software Development Group, (1995). NCSA HTTPd, Version 1.4.1, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Champaign IL 61820. http://hoohoo.ncsa.uiuc.edu/.

[Segal 1995] Segal, B.M. (1995). A Short History of Internet Protocols at CERN, http://wwwcn.cern.ch/pdp/ns/ben/TCPHIST.html.

[Sun] Sun Microsystems, Inc., Mtn. View, CA 94043-1100 USA, http://www.sun.com/products-n-solutions/hw/wstns/SS10.html

[Wall 1991] Wall, L. (1991). Practical Extraction and Report Language. Revision: 4.0.1.8, Patch level: 36.

[Wall & Schwartz 1992] Wall, L. & Schwartz, R.L. (1992). Programming Perl, Sebastopol, CA: O'Reilly & Associates, Inc. http://www.ora.com/catalog/pperl/.

[Wusage 1994] Cold Spring Harbor Labs (1994). Now from Boutell Com., Inc., http://www.boutell.com/wusage/.

[Zakon 1996] Zakon, R.H. (1996). Hobbes' Internet Timeline Version 2.4a, http://info.isoc.org/guest/zakon/Internet/History/HIT.html.

Acknowledgments

Chris Foote ** foote@chem.ucla.edu Professor Foote provided his Spring 1995 132A (Organic Chemistry) lecture as a test bed for the development stage of the VOH project. With his assistance, the details of data uploading and processing were worked out. He continues to provide us with insight into new software for use with the project.

Mary Anne Burns ** burns@chem.ucla.edu Ms. Burns is the Assistant to the Organic Division of the Department of Chemistry and Biochemistry. As well as providing technical assistance to the original development of the project, Ms. Burns also prepared some of the instructional documents.

Max Kopelevich ** mik@chem.ucla.edu Dr. Kopelevich is the Director of Computer and Network Operations for the Department of Chemistry and Biochemistry. He provided the project with the necessary computer facilities and software tools. He also donated time, experience and reference materials, which aided immeasurably in the development process.

UCLA Office of Instructional Development ** http://www.oid.ucla.edu Provided the project with a GSR IV (graduate student researchship IV) for the development of the project, and a Hewlett Packard ScanJet 4C.

Apple Computer, Inc. ** http://www.apple.com Generously donated two Power Macintosh 6100/66's and a Workgroup Server 7250.

A Java-based System for Distributed Shared Learning

Greg Millbank
Managing Director
Praxis Technical Group, Inc.

Knowing How and Knowing That

For thousands of years, the European cultural tradition has maintained a uniquely peculiar view of what it means *to learn* and what it means *to know*. Since the time of the ancient Greeks, we have nurtured a view of knowledge which firmly divides the verb *to do* from the verb *to know*. We have convinced ourselves, in spite of overwhelming evidence to the contrary, that knowing is something that happens in an individual's head, and has little or nothing to do with what the individual actually *does*. To use a distinction developed by the philosopher Gilbert Ryle, we have divorced *knowing how* from *knowing that*. We commonly believe that, for example, *knowing that* the Duke of Wellington won the battle of Waterloo is somehow a higher and more important level of knowledge than, for another example, *knowing how* to bake a cake. The great (and often misunderstood) American thinker, John Dewey, called this ancient bias the *Spectator Theory of Knowledge*, and proposed that it was the result of what he termed the *Fallacy of Intellectualism*.

Anyone who has learned to play the violin, learned karate, golf, or gardening should realize that most kinds of learning are not merely patterning of certain neural circuits in the gray matter of one's head. Learning is a whole body process, which involves the ability *to make certain moves*, only a subset of which are correct utterances, or written responses on paper. Learning is an active process involving all sorts of neural pathways, only a portion of which are in the head. Thus an ideal teaching environment has lots of opportunities for practice. It teaches *how* as well as *that*.

This fallacy has crept into our ideas about learning over the past twenty centuries. During that time it has influenced curriculum design, teaching methods, examinations, the entire school and university system, and almost all of our thinking about thinking to the point that we habitually confuse the memorization of facts with the ability to perform tasks - the ability to *recite* with the ability to *perform*. A consequence of this confusion is that we spend a lot of time and money teaching people *about* things and almost completely neglect to show them *how to perform*. In a similar way, we aim to improve performance in a job by utilizing teaching methods best suited to the memorization of facts, not the development of practices.

Individual Learning versus Collaborative Learning

A second aspect of our Western view of man and mind, is that we see human behavior as a largely individual matter. We see a person as a stand-alone entity, the act of knowing as an individual matter, and the development of a greater and more capable *Self* as the desirable end-point of *self-improvement* - the individualist's view of the individual goal of learning. However, this view of man largely ignores the fact that learning is a very *social* activity. Thinkers, such as G.H. Mead have long since described the mechanisms of the *social construction* of the self . We build our image of ourselves using the mirror of society itself. Our relations with our friends and relations are the mechanisms by means of which we construct our concept of *self*.

Furthermore, the very language in which we construct our thoughts and remember our remembrances is social in use and origin and evolves through entirely social mechanisms. The very motivation to learn, to struggle, to succeed, is also social in origin, including the socially-constructed definition of success itself. To succeed, after all, is to be accepted and regarded *by others* as a success.

The ideal learning environment is therefore a committed *group*, that is, a group of individuals with a shared passion (or at least an interest). Such a group establishes internal values which define success as movement towards attaining that common interest. This is the original idea behind schools. Such a school (as in a *school of thought*) is a community of strivers who seek a common set of goals. The members of the group reinforce each other's successful learning behavior. For example, a student violinist who must practice hours each day on her own has a very difficult struggle. But if she is a member of a *school,* she shares in the determination and enthusiasm of her fellow students. For such a group, practicing several hours each day is normal behavior.

From Good to Better

Good CBT programs emulate many of the most educative aspects of social interaction, but minimize many of the distractive effects. For example, the computer acts much like a person. It presents information, offers chances to interact, and responds to these interactions in a positive way. On the other hand, computers don't criticize, and trainees feel free to make mistakes without the prospect of public humiliation. But broader interaction - interaction with a wider group, is severely curtailed. Furthermore, interactive CBTs emulate active task-based learning by engaging the learner in simple-- although not directly related -- tasks.

Clearly, a good computer-based teaching system should maintain the successful aspects of interactive multimedia CBTs. Beyond that, however, it should add features of sociality and practicality - of group interaction and actual practice - to enhance teaching effectiveness to a new, much higher level. To accomplish these aims requires a special approach, particularly if the community of users and the teaching programs they will use are scattered across the Internet or corporate Internet, requires a special approach to software.

An ideal computerized training system would therefore retain the best of current systems, but would also encourage group learning activities, even if the members of a group were scattered over the surface of the planet. At the same time, there would be an emphasis on *doing* as an integral part of learning. Tasks would be close simulations of the actual tasks to be learned. For greater flexibility, such *active learning* opportunities could also be distributed, so that an interest group of learners could simultaneously and cooperatively work on a collection of related tasks, no matter where the tasks were located. Finally, participation in the experience should not be limited by what kind of computer the prospective learner happens to be using.

The Java Framework

Praxis Technical Group, Inc. has developed a Java-based framework technology for cooperative learning and task sharing over the Internet or Intranets.. A prototype of this system has been completed and is now undergoing testing. The framework enables complete and complex group interactions, including the sharing of data from distributed sources, and the sharing of actual experiences or simulations in real time, over the Internet or internal corporate Intranets.

Any authorized machine that has a Java enabled operating system or can run a Web browser, can participate in the framework. Within the system, users may join workgroups, or work alone, and may request a variety of types of textual, graphical and multimedia data from any number of sources in the system. Groups of users may view and discuss or manipulate the data collectively, with or without pre-defined roles, or may work alone. Information searches are carried out by an Internet-type search engine, which transparently collects information from any or all sources on the system.

The only software required on client machines is a Java run-time environment provided in most Web browsers and soon to be incorporated into most operating systems. The Praxis framework and associated client and server software can thus be automatically updated over Intranet or Internet while the system is running. Thus all software changes and upgrades may be automatically reflected on each server without the necessity of installation of new software. This dramatically reduces maintenance costs. Furthermore, any computer with appropriate security clearance, could join the system from anywhere and its user could participate in group and individual work as desired.

The advantages of the Praxis Java-based framework are many, but the main ones are those attributable to the nature of the Java language itself, including a profoundly object-oriented design. These advantages are: simplicity, easy (i.e. piece by piece) expansion/modification, ease of maintenance, and relatively low development costs. Cross platform connectivity, centralized automatic upgrades, no need for any special client software, and that it supports multi-user, multi-level, multi-role distributed cooperative workgroups are further advantages of the framework.

The concept behind the Praxis framework is *distributed communities* including *communities of communities,* their needs for information retrieval and processing, and especially the communal workgroup environment, including different levels of security clearance and different *roles.* Different roles allow different individuals with different requirements to see and deal with different subsets of information in different ways as defined by the *role type* attached to their authorization.

Since the entire framework is platform independent and completely modular, new components, including new sources, are simply added as plug-ins. Clients do not normally require any special software, and server software may be added to the system from a central location during ordinary operation. The existing framework has been designed with an emphasis on distributed servers and distributed working groups as a platform for complex multivariate simulations.

The Simulation System

The simulation system is built around certain devices and concepts:

- *Interest Groups* - groups of people who are interested in a certain aspect of a simulation or knowledge base.
- *Roles* - individuals within an interest group may interact with others in a specific way, e.g. an instructor has different interaction requirements from a trainee.
- *Scenarios* - a simulator ought to be able to emulate certain incidents or accidents - we term each emulation sequence a *scenario*
- *Scripts* - a script is a certain kind of code which controls the simulation engine in such a way as to produce a scenario. Scripts are written in a language which is both conditional and massively parallel.
- *Communications* - a communication system is provided so that individual trainees can communicate with other trainees or with *virtual people* as a part of the training exercise.
- *Virtual People* - are simulated people who take certain roles when there are no real people to fill them. Virtual People are controlled by Scripts, but can be replaced on the fly by real people.

An example:

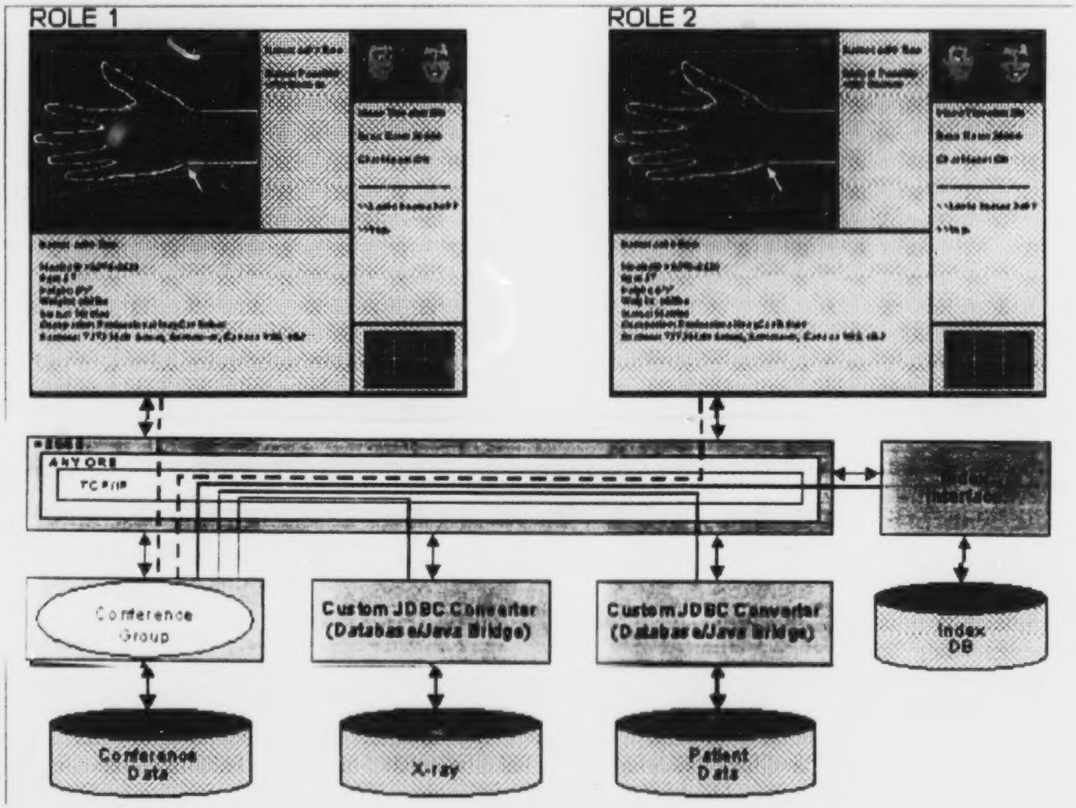
In typical operation a number of people may be involved in a simulation exercise with many dynamically interacting components. For example, consider the simulation of several ships in a Navy exercise. In the simplest form, a number of trainees would assume different shipboard roles as they interact with each other and with the simulation engine. Each trainee would be served information, including the correct visual information, according to his or her role. The steersman would get a compass and a view out the wheelhouse window (using VRML), for example, while another member of the bridge team might get a radar screen and a view through the main windows. Engine room personnel would get engine operating information and visuals, while sonar personnel would get sonar screens, etc.

Certain aspects of the simulation are controlled by a script. For example, mechanical faults may occur in the ship, other (enemy) ships or submarines may arrive on the scene (each with its own crew of real or virtual people). While scripts control certain aspects of the simulation, they also control "observations/evaluations" the simulation system makes of the trainees moves - how they exercise their roles, and so forth. Logs of the activities of the trainees are also kept.

Other ships may also appear on the screen, each powered by its own simulation engine and by its own crew (interest group) with its own collection of roles. Scripts may also interfere with the operation of these ships and watch the behavior of their crews. Because of the nature of the Java framework, other simulation engines will automatically join the simulation if they are present. The whole exercise need not be set up in advance, but can grow and adapt on the fly. These new simulation engines might be served from the same server, or served from different servers so that the total learning experience may be distributed dynamically among many computers as computing loads require.

Other Instructional Uses

The Java Framework is useful for many other purposes other than simulations. For example, a medical informatics version is under development. The following diagram illustrates its use in a medical informatics/conferencing mode. Two doctors are reviewing the same x-ray. Small video views of the participants are in the top right hand corner. Information from a variety of database sources is displayed on each screen.



The Near Future

Praxis Technical Group, Inc. specializes in producing industrial process simulations for training and engineering uses. The system will be deployed for the first of such uses within a few months. Since the framework is basically a messaging and control system, it can be adapted to a variety of other training uses with very little effort. For example, we have already combined Java-based Multimedia Programs with simulations using the framework. Multimedia computer-based training programs with, or without simulation are therefore already possible. As mentioned above, medical informatics uses are also being investigated.

Currently, the system does not work well at 28,800 baud over phone lines unless key components are downloaded to users computers first. Once major graphics files, etc. are in place on each client, bandwidth poses very little problem. However, with increased bandwidth, distributed database applications become more interesting, and the use of video conferencing improves markedly.

Praxis Technical Group is seeking interested developers in different fields to further develop the potential of this unique and powerful framework.

An Intelligent Tutoring System on the WWW Supporting Interactive Simulation Environment with a Multimedia Viewer Control Mechanism

Kiyoshi NAKABAYASHI, Mina MARUYAMA, Yoshimasa KOIKE,
Yoshimi FUKUHARA, and Yukihiro NAKAMURA

NTT Information and Communication Systems Laboratories
Tokyo 180, Japan
naka@nttaip.min.ntt.jp

Abstract: This paper describes the new features of an intelligent tutoring system(ITS) called CALAT. The architecture of CALAT is an extension of conventional WWW system, consisting of ITS kernel on the server side and multimedia viewer on the client side. A viewer control mechanism is newly developed to achieve both educationally effective multimedia presentation and quick system response time even over the low speed network. As a application of the viewer control mechanism, an interactive simulation environment is implemented. This environment let the student interact with a target(simulated) system presented by the multimedia viewer under the control of a state transition machine(STM) running on the CALAT server representing the internal behavior of the target system. These new features make it possible to build an interactive ITS environment available over the network with powerful and effective data presentation capability.

1 Introduction

Since its capability as a globally distributed hypermedia is so attractive, the World-Wide Web(WWW) [Berners-Lee, 1994] has been widely accepted by not only computer scientists but the people in the various fields. One of the important application of the WWW is for education. There have been many proposals to use the WWW in a educational environment [Kay & Kummerfeld, 1994] [Ibrahim 95] [Schwarz et al, 1996].

We have proposed CALAT [Nakabayashi et al., 1995], an intelligent tutoring system(ITS) on the WWW. The system consists of an WWW server integrated with a tutoring system called CAIRNEY [Fukuhara & Kiyama, 1993] [Fukuhara et al., 1995] which has been developed as a standalone ITS, and WWW clients equipped with a multimedia scene viewer. The CALAT system features the mechanism to identify the user on the client side from the sever side over the stateless protocol of the WWW. This mechanism makes it possible to implement substantial ITS function to present appropriate courseware pages by taking into account of the user's answers to the exercises, progress speed, and other characteristics observed from the interaction history.

Although the first version of CALAT is actually implemented and open to the public on the Internet, there are several limitations and problems:

- Size of courseware page consisting of animation and audio data is too large to be transfered over the low speed network. Sometimes it takes more than a few minutes to display the page, making learning substantially impossible.
- Learning style is limited to explanation and exercise based. There is no support for interactive simulation which, if available, could provide a very effective tutoring environment by the combination with the explanation-based tutoring.
- It is not possible to utilize existing HTML documents and other type of multimedia data distributed over the Internet.

The design goal of new CALAT system is to provide solutions for these issues. This paper concentrates on the first two topics. The third topic is dealt with in another paper [Maruyama et al., 1996]. A viewer control mechanism is newly developed for enabling fast system response possible even for the courseware page of large data size transmitted over low speed network. The mechanism also provides a sophisticated viewer control facility important to realize effective tutoring. Exploiting the viewer control mechanism, an interactive simulation environment is implemented. In this environment, the internal behavior of the target(simulated) system is represented by a state transition machine(STM) running on the CALAT server. The STM fully takes control of the multimedia viewer on the client side by means of the viewer control mechanism so that the student can experience how the target system responds to his/her operation.

The following sections, the design and implementation of the viewer control mechanism and the interactive simulation environment are described.

2 Viewer Control Mechanism

2.1 Design Issues

Multimedia data presentation is quite important for the CAI system. There are several crucial issues to support effective multimedia presentation in the WWW environment.

One issue is the data transfer speed over the network. The amount of multimedia data can be so large that the system's response to the user may be quite slow. In the case of CALAT, one standard courseware page consists of about 10KByte animation data and a few hundred KBytes audio data replayed synchronously. Since these data are packed in one huge archive file [Nakabayashi et al., 1995], the users on the ISDN or telephone line should wait for a few minutes until the requested page is downloaded.

Moreover, current implementation can not take advantage of the data caching mechanism equipped in the WWW proxy server or browser. This is because the each courseware page is selected by the CALAT server responding to the ``same'' URL request from the client saying ``send me the next page'' or ``give me one more hint''. This behavior is essential to implement the adaptive tutoring capability of CALAT with which the courseware pages are dynamically selected based on the student status. However, The proxy server or browser cannot cache the previously downloaded pages since the URL's corresponding to them are all identical and not represent specific data.

Another issue is the viewer control from the server. In the conventional WWW scheme, it is very difficult to make an effective presentation using multiple client window, since there are no means for the server to take control of the external viewers (or helper applications) spawned by the WWW browser. It is important for the effective tutoring since since sometimes it is good for the student to achieve a good understanding by looking at multiple images at a time or sometimes concentration on one scene may result in much progress.

2.2 Implementation

The previous version of CALAT has the mechanism to partially solve the second issue mentioned in the previous section. The new version provides an extended scheme which is designed to overcome the above two issues without modifying the conventional WWW browser and protocol.

The key component of the extended scheme is a control script sent from the CALAT server to the client as the response to the student's courseware page request. On receiving the URL from the client saying ``send me the next page'' or ``give me one more hint'', the CALAT server assembles the control script taking into account of the student status. The WWW browser on the client receives the script and invoke a viewer controller program which interpret the script. The script contains the following information for presenting the courseware page:

- The type and the identifier of the viewer to be used to display the page. If the viewer of the specified type and identifier is not running on the client, new one will be invoked. If it is already running, its current display data will be thrown away and the new data will be loaded. It is also possible to close the specified viewer if necessary.

- A list of possibly multiple URL's specifying the multimedia data composing the page. The viewer controller requests these URL's by means of the WWW browser's remote control facility such as CCI [NCSA, 1995] for Mosaic or NCAPIS [Netscape, 1995] for Netscape. The received multimedia data will be passed to the viewer specified in the first item.

With the capability described in the first item, it is possible for the server to take control of multiple viewers on the client. For example, the new version of CALAT is able to open a new viewer for the detailed explanation of a certain keyword, or to automatically close every viewer except the browser when the student comes to an exercise page [Fig.1].

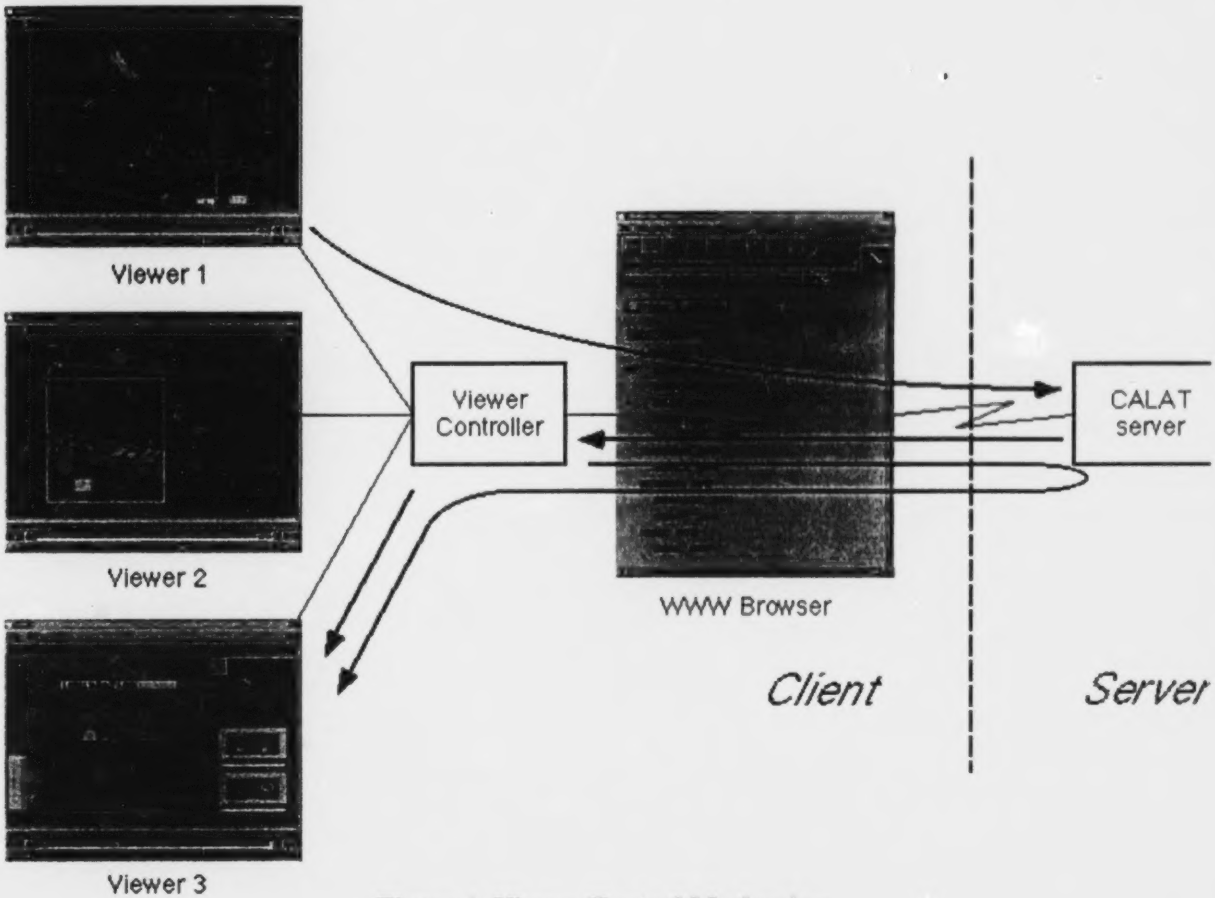


Figure 1. Viewer Control Mechanism

The capability of the second item allows to construct a courseware page consisting of multiple multimedia data. One application of this feature is the interactive simulation described in the following section, in which it is required to display image and sound simultaneously as the response to the user's single action.

Another application of this feature is the ``pipelined processing'' mode to make the system response faster[Fig.2]. In the pipeline processing mode, a few hundred KBytes audio data belonging to a CALAT courseware page will be splitted into the small (usually some ten KBytes) chunks. URL's of these audio data chunks together with the animation data will be listed in the control script. The viewer controller accesses these URL's one by one and forward the received data to the viewer. As soon as the viewer receives the animation data and the first chunk of the audio data, it starts to replay them. At the same time the viewer controller keeps accessing the succeeding audio data. Thus the data transfer over the network and replay on the client are performed in a pipelined fashion. With this pipeline processing mode, it is possible to drastically shorten the system response time especially in the case that the network speed is slow and/or the amount of multimedia data in a page is large.

Moreover animation and audio data can be cashed by the WWW proxy server or the browser, since they are accessed with its specific URL's with which the WWW proxy server or the browser distinguish each of them. Thus when the student go back to the previous page, the control script for the previous page comes from the server but the multimedia data composing the page may come from the cache with high probability. This results much more faster response.

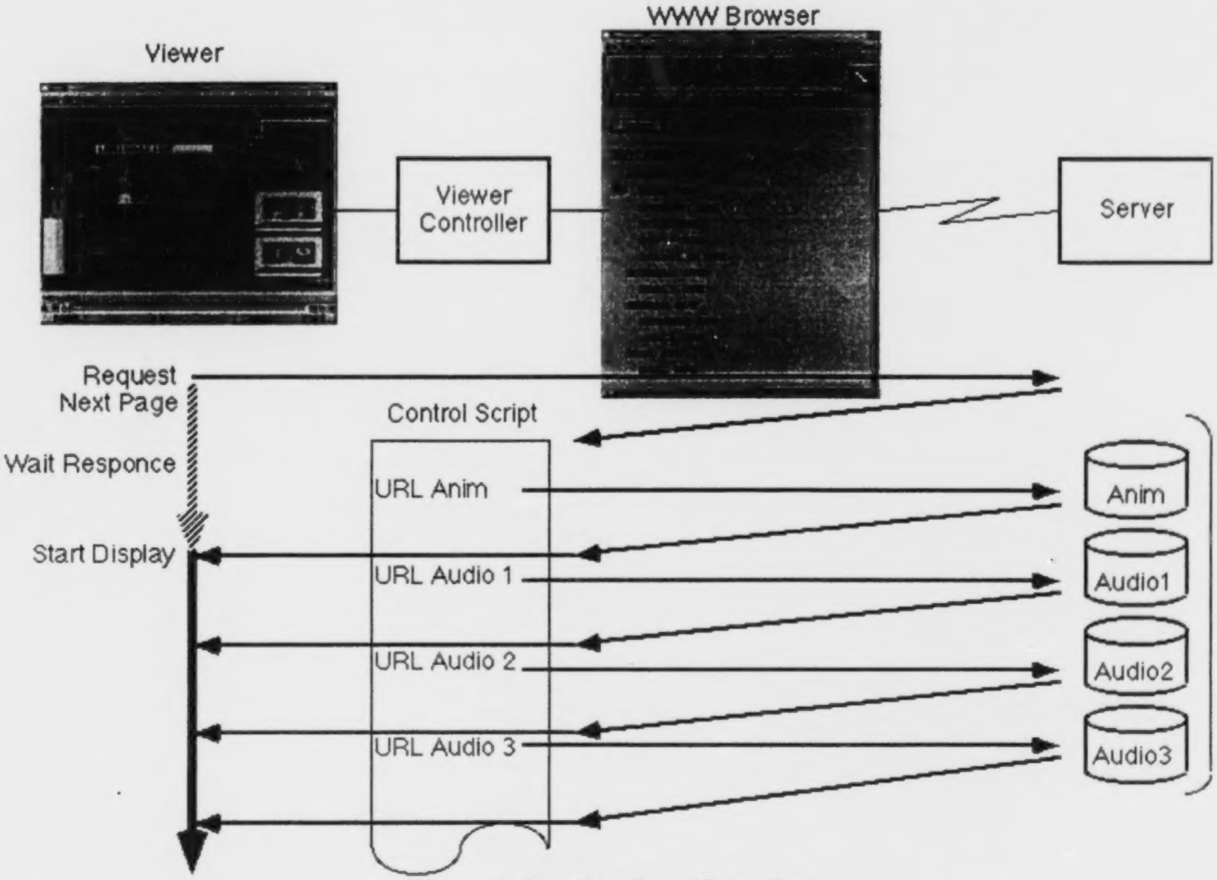


Figure 2. Pipeline Data Transfer

It should be noted that the proposed control script scheme works completely in the conventional WWW environment. Since the WWW browser deals with the control script in a same way as a usual multimedia data displayed using the helper application, the viewer controller program is just registered as a helper application to be spawned when the browser receives the control script. Also the remote control facility used to access URL data by the viewer controller is supported in the commonly used WWW browser such as Mosaic and Netscape. Thus using the control script poses no special modification of the WWW browser or protocol.

3 Interactive Simulation

3.1 Background: CAIRNEY Simulation

An interactive simulation is one of the most important facility to make effective tutoring possible.

CAIRNEY [Kiyama & Fukuhara, 1993], the standalone ITS on which CALAT is based, has the simulation capability mainly for the training of operation procedures to deal with the telecommunication network equipments.

In the CAIRNEY simulation environment, the behavior of the simulated(or target) network equipment is described as a state transition model(STM). The STM consists of several states, the input event which triggers the state transition, and the output actions taking place when certain state transition occurs. The input event and output action of the STM are linked to the user interface of the CAIRNEY simulation environment. The user interface window consists of simple GUI(graphical user interface) objects(widgets) such as buttons or key-input fields, and multimedia outputs such as animated images or sounds. The user's click or type on the widgets causes the event input to the STM, and the output actions from the STM change the color of the widgets, redraw the window image or make sound output.

In the old version of CALAT, above mentioned simulation facility is not implemented because the conventional WWW framework is not adequate to support the required user interface. Standard WWW browsers provide only anchor, form input and clickable map as the ways which user can interact with. There are no concept of ``widget''. It is also difficult or impossible to make effective output such as partially modifying the previously displayed image, or redrawing the image while playing audio simultaneously. (Even recent technology such as VRML does not support ``widget''. Java could be a solution but not very popular at this point of time.)

3.2 Design and Implementation

There are two major design issues to implement CAIRNEY-style simulation on CALAT environment. One issue is how to divide the function composing the entire simulation between the server and client. Another issue is how to implement ``widget'' in the WWW client environment.

Two main alternatives can be considered for the first issue. One is to implement both STM and GUI on the client. It gives maximum response performance for the user's mouse click or key input. The drawback is that we have to prepare STM for every client platform. Moreover the CAI kernel on the server and STM are rather isolated, making it difficult to give pedagogical intervention from CAI kernel. The other alternative is to implement the STM on the server and the GUI on the client. In this alternative, the response speed may be slower than the first one. However it is much much easier to make the simulation environment available on the multiple client platform. Since there are actual requirements to use CALAT system with several client machine, the second alternative is taken for less development effort.

For the second issue on how to implement interactive widgets, we fully take advantage of viewer control mechanism described in the previous section. We developed the multimedia viewer which works together with the viewer controller. The multimedia viewer is able to display the animation/audio data sent from the viewer controller as well as to manage widgets on the window. Mouse click or key input event on the widgets are sent to the viewer controller, then the viewer controller translates it to the URL form and send it to the CALAT server. This will trigger the state transition of the STM on the server. As the result of the state transition, the server generates the viewer control script and send it back to the client making changes on the window and/or playing audio. With this mechanism, CALAT provides an interactive simulation environment available on the network-based server/client system of the WWW. Example of the interactive simulation is shown in [Fig.3]

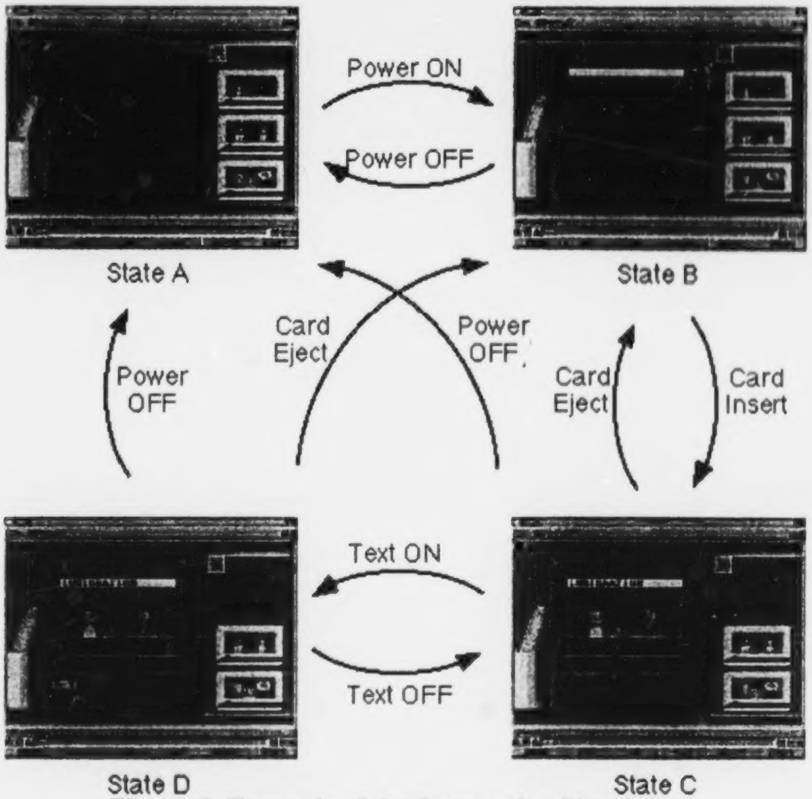


Figure 3. Example of the Interactive Simulation

4 Conclusion

New version of CALAT, an ITS on the WWW is described. It is equipped with several novel features including a viewer control mechanism and an interactive simulation facility. The viewer control mechanism make is possible to achieve a fast response speed even over low speed network as well as to provide sophisticated multimedia presentation. The simulation environment is implemented by fully exploiting the viewer control mechanism. It let the student interact with a target system managed by the STM on the CALAT server over the network with powerful and effective multimedia presentation.

References

[Berners-Lee, 1994]
Berners-Lee, T., et. al. (1994). The World-Wide Web. *Communications of the ACM,* 37(8), 76--82.
[Schwarz et al, 1996]
Schwarz, E., Brusilovsky, P., & Weber, G. (1996) World-wide intelligent textbooks. *Proceedings of the the ED-MEDIA 96.*
[Fukuhara et al., 1995]
Fukuhara, Y., Kimura, F., Kohama, C., & Nakamura, Y. (1995). A Knowledge-based Educational Environment Integrating Conceptual Knowledge and Procedural Knowledge in Telecommunication Service Field. *Proceedings of the ED-MEDIA 95,* 229--234.
[Fukuhara & Kiyama, 1993]
Fukuhara, Y., & Kiyama, M. (1993). A Knowledge-based Educational Environment for Real Business Settings. *Proceedings of the World Conference on Artificial Intelligence in Education.*
[Ibrahim 95]
Ibrahim, B. and Franklin, S.D. (1995). Advanced Educational Uses of the World-Wide Web,

Proceedings of the Third International World Wide Web Conference '95: Technology, Tools and Applications

[Kay & Kummerfeld, 1994]

Kay, J. & Kummerfeld, R.J., (1994). An Individualized course for the C programming language. *Proceedings of the Second WWW Conference.*

[Kiyama & Fukuhara, 1993]

Kiyama, M. & Fukuhara, Y., (1993). An Authoring System for Creating a Practice Environment in the Network Service Field. *Proceedings of the International Conference on Intelligent Computer-Aided Training and Virtual Environment Technology.*

[Nakabayashi et al., 1995]

Nakabayashi, K., Koike, Y., Maruyama, M., Touhei, H., Ishiuchi, S., & Fukuhara, Y. (1995). A distributed Intelligent-CAI System on the World-Wide Web. *Proceedings of the ICCE 95,* 214--221.

[NCSA, 1995]

National Center for Supercomputing Applications (1995). NCSA Mosaic Common Client Interface. http://www.ncsa.uiuc.edu/SDG/Software/XMosaic/CCI/cci-spec.html.

[Netscape, 1995]

Netscape Communications Corporation (1995). Standards Documentation. http://home.mcom.com/newsref/std/.

[Maruyama et al., 1996]

Maruyama, M., Nakabayashi, K., Koike, Y., Fukuhara, Y. & Nakamura, Y.(1996). An Approach to Implement Adaptive Hypermedia for an Intelligent Tutoring System on the World-Wide Web. *Proceedings of the ED-MEDIA 96,* 373.

Designing Digital Libraries for Post-literate Patrons

Peter J. Nürnberg, Erich R. Schneider, John J. Leggett
Center for the Study of Digital Libraries, Texas A&M University, USA
email: {pnuern, erich, leggett}@csdl.tamu.edu

Abstract: Many Web and Internet technologies have traditionally been used to serve information stores across machines and between people. There has been a great deal of recent interest in using these information services to support digital libraries. Digital libraries is an interdisciplinary research effort that must synthesize existing research from highly disparate fields. This paper examines two such contributing fields - information systems and orality-literary studies - and applies them to a particular digital library domain, botanical taxonomic work. In trying to build digital libraries for botanical taxonimists, we show how two widely differing fields each can provide part of a solution neither can provide alone.

0. Description of the Paper

Many Web and Internet technologies have traditionally been used to serve information stores across machines and between people. There has been a great deal of recent interest in using these information services to support digital libraries. Research in digital libraries, as any interdisciplinary endeavor, is confounded by the fact that one must consider and synthesize from several fields, including ones that are perhaps unfamiliar. This paper seeks to tie together two very different fields - information systems and orality-literacy studies - that each have something to offer the digital library designer. The authors have chosen an unconventional format for presenting this material. The paper contains two threads for the two fields it draws upon. Some sections belong in only one thread, while others belong in both. The paper can be read in different ways, but most people will find it easiest to read the thread with which they are more familiar first, in order to contextualize the material, and then delve into the other thread. Figure 1 below illustrates the organization of the paper. (Note: the information systems thread should be read from the left column and the orality-literacy thread from the right).

Figure 1: Organization of the paper

0. Description of the paper	
IS1. Introduction	OL1. Introduction
IS2. HOSS Architecture	OL2. Orality, Literacy, and Hyperliteracy
3. Botanical Taxonomic Scholarship	
IS4. Technology Applications	OL4. Hyperliterate Work Practices
5. Conclusions	

IS1. Introduction

For many reasons, archaic work practices of varying "inappropriateness" to modern scholarship linger on despite their known flaws. In information-intensive fields, considerable support for the development of new work practices can be provided by digital libraries and the

OL1. Introduction

For many reasons, archaic work practices of varying "inappropriateness" to modern scholarship linger on despite their known flaws. In information-intensive fields, the derivation of possible new work practices can be suggested by differentiating those aspects of current practice that are archetypic to the problem addressed from those artifactual to the technologies currently employed. In particular, orality-literacy technologies underlying them. In particular, advanced distributed, computationally-oriented hypermedia systems, with their capability to support more fluid information structures, have often been proposed for use in fields where the mutable cognitive artifacts that scholars employ are known to be poorly reflected in the static artifacts produced by pre-electronic work practices for pre-electronic distribution methods.

IS2. HOSS Architecture

HOSS is a computationally-oriented hypermedia system [Nürnberg et al. 1996]. It consists of a hyperbase layer, a structure processing layer, a metadata manager layer, and an application layer. Each of these will be briefly described below.

The main difference between HOSS and other hypermedia systems is that HOSS is an entire operating environment. It provides file system, memory management, and scheduling features. Other operating system functionality is provided by a SunOS 5.4 kernel. HOSS is best thought of as a hypermedia-aware operating system. An immediate result of this is that HOSS, as any operating system, admits an open set of application processes. Furthermore, just as all applications in a real-time operating system may take advantage of real-time awareness on the part of the operating system, all HOSS applications have immediate access to hypermedia functionality. The functionality of the hyperbase and (open) structure processing layer is available to all HOSS processes.

studies are here proposed for this pupose in fields where the mutable cognitive artifacts that scholars employ are known to be poorly reflected in the static artifacts produced by pre-electronic work practices for pre-electronic distribution methods.

OL2. Orality, Literacy, and Hyperliteracy

Since the 1960s an interdisciplinary research area within the humanities known as orality-literacy studies has existed, concerned with differences in the modes of thought and expression exhibited by individuals in cultural situations which exhibit *primary orality* (where writing is not used as an adjunct to thought and memory) and those exhibiting *pervasive literacy* (where it has become indispensable for those activities).

OL2.1 Orality and Literacy

A seminal work in orality-literacy studies is *Preface to Plato* by classicist Eric Havelock [1963], whose starting point is Plato's attack on poetry in the *Republic* [Waterfield 1993]. Plato's proposal that poetry be banned from his ideal state, because it degraded the intellect, is found odd by many modern students of Plato. Havelock sets out to examine what this apparent oddity in the philosopher's thought implies about the cultural situation of Plato's Greece.

Havelock contends the extensive ground of common knowledge and worldviews required by classical Greek culture was encoded in the great poems of the time, most notably Homer's epics. To the ancient Greeks, these were a "tribal encyclopedia" of cultural ways and norms. Poetry was also well-suited to the problems of information storage in a non-literate culture, namely retention in living memory and content-preserving transmission [Havelock 1963]. In essence, recitation of the epics was able to induce in reciters and listeners an almost

IS2.1 Hyperbase Layer

A HOSS hyperbase is a process with two threads: a Versioned Object Manager (VOM) and an Association Set Manager (ASM). The VOM acts as a client of some Storage Manager (SM) that exists outside of the hyperbase. The VOM serves simple object and composite object abstractions and provides full versioning support for both [Hicks 1993]. The ASM is implemented as a client of the VOM, mapping the VOM abstractions to structural entity abstractions called associations and association sets [Leggett and Schnase 1994; Schnase 1992]. Because the ASM is a client of the VOM, it inherits versioning support for its abstractions as well.

A HOSS hyperbase is conceptually similar to other hyperbase systems [Leggett and Schnase 1994; Schnase 1992; Shackelford et al. 1993; Schütt and Streitz 1990; Wiil 1993].

IS2.2 Structure Processing Layer

HOSS allows an open set of structure processors called Sprocs. All Sprocs are clients of the ASM. The difference between Sprocs lies in the kinds of structure they manipulate. A key aspect to HOSS Sprocs is that they abstract behavior from structure [Nürnberg et al. 1996].

One example of a HOSS Sproc is the Link Services Manager (LSM). The LSM manages ^{3}traditional^2 hypermedia structure - namely, inter-application linking structure. It provides functions to create, navigate, manipulate, and destroy structure between application data. In the case of the LSM, behaviors correspond to the semantics of particular navigational structure traversals.

The Taxon Manager (TaxMan) provides a second example of a HOSS Sproc. TaxMan acts as a client of both the VOM and the ASM, and serves taxonomic structural abstractions. These taxonomic abstractions are widely

hypnotic state that assisted correct remembrance. It also encoded cultural knowledge situationally. Both of these were anathema to Plato, who was promoting reflective thought on the nature of abstracts. Plato's literacy allowed him to encode knowledge externally as a thing "in itself" and allowed him to examine concepts and their abstract structures without forgetting them. Thus, Havelock concludes, arises Plato's excoriation of poetry as education method, as inhibitor of abstract speculation on the nature of the true, good, and beautiful. For our purposes, we note that Havelock showed the consideration of ideas as eternal "things in themselves" is an artifact of literacy, not an archetypic aspect of thought.

Table 1: Examples of Differences Between Orality and Literacy.

	Orality	**Literacy**
Ideas as... [Havelock 1963]	properties of concrete situations	abstract and eternal "things in themselves"
Socially relevant truths as... [Ong 1982]	mutable objects	fixed objects
Language use as... [Ong 1982]	requiring consideration of situation	manipulation of abstract placeholders

Among other artifactual properties of literacy (examined in another seminal work of the field, Walter Ong's *Orality and Literacy* [Ong 1982]) is the notion of written truth as permanent truth. Today, it is common for material to be written down and remain unchanged for extended periods of time. If that material had some veracity when it was recorded, we tend to regard its "truth" as a permanent property that can be redemonstrated at any time. This is not the case with orally transmitted knowledge, which cannot be "recorded" except in living memory. As a result, material for which there is no call is forgotten, and changes to the material that give advantage will occur. Revisionism is reality in primary oral cultures; the beliefs that the written retains its truth for all time and that, by extension, publication implies truth are artifacts of literacy.

OL2.2 Hyperliteracy

Many believe that we are entering an era where electronic tools for storing and manipulating taxonomic abstractions are widely applicable. For example, botanical taxonomists use abstractions such as family, genus, species, etc. to classify plant specimens. Also, linguists develop linguistic taxonomies to represent the developmental histories of different languages.

Additionally, the TaxMan provides a number of standard computations over taxonomic structures (i.e. behaviors). Some examples of these behaviors include structure querying (e.g. find all family taxa that contain four genera with only one species each) and ³structure collapsing² (e.g. collapsing species, subspecies, section, etc. into the genera taxa and transferring the associations between specimen data and these collapsed taxonomic levels to the genera.)

IS2.3 Metadata Manager Layer

Metadata managers are system processes that primarily serve abstractions to other system processes. They build the abstractions they serve from abstractions served by other metadata managers, Sprocs, and hyperbases. Metadata managers can be viewed as abstract data types, exporting data and functional abstractions.

IS2.4 Application Layer

Application processes are user processes familiar from conventional operating systems. The nature of these processes is open. One example of an application that has been built is a WWW Common Gateway Interface (CGI) [Berners-Lee et al. 1992] program that acts as a client to the TaxMan, allowing queries to be made over a taxonomic space, displaying the results, and allowing users to annotate the records displayed in answer to the query. Another example is a Motif/X [Nye 1988; Young 1990] client that allows graphic editing and manipulating of taxa.

IS2.5 Other Tools

electronic tools for storing and manipulating information will be considered indispensable for everyday thinking and remembering. Douglas Engelbart [1963] expressed this belief when he described a "certain progression of our intellectual capabilities", from *concept manipulation* (manipulation of concepts in the mind alone) to *symbol manipulation* (expression of concepts through language) to *manual external symbol manipulation* (manipulating linguistic symbols using writing) and finally to *automated external symbol manipulation* (manipulation of symbols using computers). Engelbart's second stage corresponds with the concept of "primary orality", and his third stage with "pervasive literacy". We extend the concept of orality and literacy by positing a new property of culture, *pervasive hyperliteracy* or simply *hyperliteracy*, corresponding to Engelbart's fourth stage.

Why posit hyperliteracy? If we are indeed entering an era where automated external symbol manipulation tools have become prerequisites of serious thought, then the designers of such tools should be interested in which aspects of thought are intrinsic to language-using human beings and which aspects are products of the use of non-electronic writing, since some of the latter may decrease in strength or disappear altogether in the residents of this new era. As can be seen from the above, these artifactual properties are not trivial, and they are precisely the concern of orality-literacy studies.

A number of tools have been built for application, metadata manager, and Sproc construction [Nürnberg 1994]. The HCMT and HPMT toolkits provide certain process model and inter-process communication primitives. A tool called the PDC allows quick construction of servers by generating the necessary protocol libraries from high-level protocol specifications.

3. Botanical Taxonomic Scholarship

A curious aspect of some scholarly work practices is that often, these practices are known to depend on false assumptions or over-simplifications of a problem. In some cases, such as in certain economic models, these false assumptions are taken as reasonable because they produce good results and make the models tractable.

In other cases, however, these false assumptions are simply products of tradition, based in part on artifacts of old technology and literate mindsets. We take as one very specific example our experiences with botanical taxonomists. For several years, we have worked together with botanists to build a digital library of herbarium collection data. We have been able to observe several common current work practices that have changed as our botanist colleagues both gain access to new technology and re-evaluate those parts of their old technology that dictated how they did their jobs. As a particularly good example of a current work practice dictated by current technology, consider that there are journals that use taxonomies that everyone (including the journal editors!) acknowledges are outdated. The editors of the journal, however, are reluctant to correct the errors in this standard taxonomy, partly because the fixes are not universally agreed upon, but also because changing the taxonomy now would "invalidate" articles just published. The current common practice, then, is for researchers to carry out their work using a more realistic taxonomy, and then literally "uncorrect" their terms to match the journal standard.

For reference, the object of taxonomic classification is the *taxonomy*, which consist of *taxa*, which themselves consist of other taxa or *specimens*. Taxa are composed in a hierarchic fashion. Taxa at different levels in the tree have different names, such as family, genus, species, etc. We briefly describe three interesting problems we observed the taxonomists encounter in their current work practices.

Different groups of taxonomists produce different taxonomies, even if the specimen set examined is identical. Groups in which particular specialists work on a given taxon may show more detail in the expansion of that taxon, or different groups may use different measures of similarity when composing taxa, weighting various kinds of evidence differently. It seems contradictory to have multiple solutions to a classification problem.

Separate taxonomic groups produce separate taxonomies, which are then identified by the groups that produced them. This is despite the fact that it may always be used in conjunction with other taxonomies, or that it is based on the prevailing attitudes in the community. It seems contradictory that a communally defined, communally used product is identified with a small set of taxonomists.

The products of the work are often taxonomies, not simply revisions to existing taxonomies. Whether updates or new full revisions, the products are viewed as a closed, well-defined entities, representing an opinion of a group at some time. However, new evidence, new analysis methods, and new interpretations are constantly being introduced. It seems contradictory to produce a well-defined, static analysis of an ill-defined, dynamic phenomenon.

IS4. Technology Applications

Addressing the three examples of seeming contradictions in current work practices requires different supporting technologies than those present in the physical library. What is required here are new digital library elements and tools, not derived from physical antecedents. Of course, it is impossible to say what all of these technologies will be. This section outlines some possible technologies to begin to address these issues.

IS4.1 Single/Multiple Taxonomies

Two important capabilities that helps address single/multiple taxonomies problems are structure management and versioning. Hypertext structure management abstracts the structure over objects from the objects themselves. Oftentimes, this takes the form of abstracting traversal or navigational structure from data to be navigated. However, the principle of structure abstraction can be applied to *any* realm in which multiple structures may be applied to a given data set. This is precisely the case in taxonomic work. Different taxonomies (structures) are built over the same specimen (data) set. Because the TaxMan inherits the structure management abstractions of HOSS, including contexts (sets of structure elements and their associated behavior processes), it can use these contexts to partition the taxonomic data into consistent taxa sets.

Because the TaxMan is implemented on top of HOSS, it inherits the versioning support for both data and structural objects therein. This provides a natural way to model difference over time in a given taxonomy, as well as differences with respect to authority in the same time frame. Additionally, changes in the analysis of specimens (perhaps the addition of new pictures or new genetic information) can be added to the data set by versioning the appropriate specimen data object, thereby not invalidating taxonomies

OL4. Hyperliterate Work Practices

Addressing the three examples of seeming contradictions in current work practices requires different artifacts than those present in the physical library with its literate artifacts. What is required here are new digital library elements and tools, not derived from physical antecedents [Nürnberg et al. 1995]. Of course, it is impossible to say what all of these artifacts will be. This section outlines some possible artifacts to begin to address these contradictions.

OL4.1 Single/Multiple Taxonomies

One artifact of literacy is the notion of single-valued, static truths [Ong 1982]. The work practice of developing and publishing taxonomies separately from one another is a particular instantiation of this artifact. The product of this work is a taxonomy, a "taxonomic fact" or truth, presented and interpreted as such. However, the notion of truth is changing from the literate view of static and single-valued to the hyperliterate view of dynamic and multi-valued. Consider the Guides project approach to teaching history in which various persona contextualize history from a particular point of view [Solomon et al. 1989]. The "truth" of the matter is a space, in which various points of view are represented. This contrasts sharply with the notion of the authority of the book as conveyor of a single, coherent message as in the literate world [Chartier 1994]. Perhaps instead of viewing the primary goal of a taxonomist as the generation of a new taxonomy, which then must be related to previous and competing taxonomies by the consumer, the product may be viewed as a change to the existing body of knowledge. In fact, in essence, taxonomists *do* view the *purpose* of their work in this way, but the actual product of their work, the printed taxonomy, is only a means to this end. Reconciliation and contextualization is the responsibility of the consumer.

based on the older version of the object.

IS4.2 Ownership of Taxonomies

One important capability that helps address ownership of taxonomies problems is annotation support. An important aspect of maintaining and using community objects is annotating and sharing annotations over community objects. Such annotations can be used to judge the communal level of acceptance of a part of the community body of knowledge or other particularly noteworthy aspects. Moderately sophisticated access control, search facilities and filtering mechanisms over the annotation space should be provided. We have developed a HOSS Sproc named AnnoMan, which models sets of annotations as structure contexts and provides these features. Modeling annotations as structure in a hyperbase is straightforward - different structural elements (annotations) are laid over existing taxa and specimens (data), grouped into contexts, and managed by existing hyperbase software that can provide access control.

IS4.3 Definition of Taxonomies

Another important capability that helps address definition of taxonomies problems is computation over hypermedia structure. The nature of the information in taxonomic research may be open in the sense that the boundaries around it may be hard to define, especially outside of a particular context. However, dealing with documents that exhibit no sense of closure at all can be disorienting as well. What is needed is a way in which the open space can be viewed as only "partially" open - that is, enforcing some sort of boundaries appropriate in a context, but allowing these boundaries to be crossed or recomputed. One way in which to do this is to take advantage of computation over structure which dynamically generates closed sets of structure appropriate for a particular use.

OL4.2 Ownership of Taxonomies

Literacy promotes the concept of idea ownership by the individual, even when the idea represents a communally held truth. In this case, taxonomies are identified with their producers or publishers. There is no way to recognize the contextualization of a taxonomy in itself. Howevere, the notion of authorship is changing from owner of a document and by extension its ideas to recorder of ideas that are the product of several people, past and present. Consider an analogy from the business world - the growing role of the analyst [Reich 1991]. The analyst provides a filtering or ordering function for data that is oftentimes already available. Many new companies focus no longer in the production of information, but its compilation. This reflects a situation in which the problem of information is what to do with the overabundance of it (the "information explosion"), and not how to find and retrieve data [Chartier 1994].

OL4.3 Definition of Taxonomies

One artifact of literacy is closure of ideas. The product of taxonomic work is a well-defined, discrete entity. Products no longer must be closed. They may exist as changing entities over time, with poorly defined borders. Consider Web sites with links to many other sites. These sites have no closure per se. Where one chooses to draw boundaries is contextually and individually defined. This is in opposition to the closure engendered by books and other written entities [Chartier 1994]. As above, one new possibility is a communally maintained set of taxa, with various notes, modifications, and addenda separately maintained over these taxa. The boundaries of the communal knowledge could only be determined by a given consumer at a given moment.

5. Conclusions

The information systems thread of this paper asserted the existence of new work practices in botanical taxonomic scholarship enabled by new technologies. The new work practices, however, were assumed to arise spotaneously due to problems found in current work practices.

The orality-literacy thread of this paper motivated why certain new work practices might arise in botanical taxonomic scholarship, but did not offer any particular ways to cope with them.

The digital library will have to support the new work practices of people. The changes in such practices must be identified. We extended orality-literacy to hyperliteracy in an attempt to characterize the changes. The new practices will have to be supported by new technologies. We showed systems and tools able to support the needs of one particular research community. The threads in this paper, therefore, must rely upon one another, one for motivation, the other for prototypic solutions. We see this as a microcosm of the digital libraries research field - a field in which results from many different and dissimilar areas will need to be synthesized to produce the research necessary to redesign the tools with which people think.

6. References

[Berners-Lee et al. 1992] Berners-Lee, T. J., Cailliau R., Groff, J. F., Pollermann B. (1992). World-Wide Web: The information universe. *Electronic Networking: Research, Applications and Policy* 2 (1), 52-58.

[Chartier 1994] Chartier, R. (1994). *The Order of Books: Readers, Authors, and Libraries in Europe Between the Fourteenth and Eighteenth Centuries.* Stanford, CA: Stanford University Press.

[Engelbart and English 1968] Engelbart, D. C., and English, W. (1968). A research center for augmenting human intellect. *AFIPS Conference Proceedings, 1968 Fall Joint Computer Conference*, San Francisco, CA.

[Havelock 1963]. Havelock, E. (1963). *Preface to Plato*. Cambridge, MA: Belknap Press.

[Hicks 1993] Hicks, D. (1993). A Version Control Architecture for Advanced Hypermedia Environments. Ph.D. dissertation, Department of Computer Science, Texas A&M University.

[Leggett and Schnase 1994] Leggett, J., and Schnase, J. (1994). Viewing Dexter with open eyes. *Communications of the ACM*, 37 (2), 76-86.

[Nürnberg 1994] Nürnberg, P. (1994). Implications of an open, extensible, and distributed hypermedia information system architecture for inter-process communication subsystem design. M.S. thesis, Department of Computer Science, Texas A&M University.

[Nürnberg et al. 1995] Nürnberg, P., Furuta R., Leggett, J., Marshall, C., and Shipman, F. (1995). Digital Libraries: Issues and Architectures. *Proceedings of the Digital Libraries '95 Conference*, Austin, TX.

[Nürnberg et al. 1996] Nürnberg, P., Leggett, J., Schneider, E., and Schnase, J. (1996). Hypermedia Operating Systems: A New Paradigm for Computing. *Proceedings of the Hypertext '96 Conference*, Bethesda, MD.

[Nye 1988] Nye, A. (1988). *Xlib Programming Manual for Version 11*. Sebastopol, CA: O'Reilly and Associates Inc.

[Ong 192] Ong, W. (1982). *Orality and Literacy: the Technologizing of the Word*. New York: Methuen.

[Reich 1991] Reich, R. (1991). *The Work of Nations: Preparing Ousrselves for 21st Century Capitalism*. New York: A. A. Knopf.

[Schnase et al. 1993] Schnase, J., Leggett, J., Hicks, D., Nürnberg, P. and Sanchez, J. 1993. HB1:

Design and implementation of a hyperbase management system. *Electronic Publishing: Origination, Dissemination and Design*, 6 (1), 35-63.

[Schütt and Streitz 1990] Schütt, H., and Streitz, N. (1990). HyperBase: a hypermedia engine based on a relational database management system. *Proceedings of ECHT '90*, Versailles, France.

[Shakelford et al. 1993] Shackelford, D., Smith, J., and Smith, F. (1993). The architecture and implementation of a distributed hypermedia storage system. *Proceedings of the Fifth ACM Conference on Hypertext (Hypertext '93)*, Seattle, WA.

[Solomon et al. 1989] Solomon, G., Oren, T., Kreitman, K. (1989). Using Guides to explore multimedia databases. *Proceedings of 22nd Hawaii International Conference on System Sciences*, Kailua-Kona, HI.

[Waterfield 1993] Waterfield, R. (trans.) (1993). *The Republic of Plato*. New York: Oxford University Press.

[Wiil 1993] Wiil, U. (1993). Experiences with HyperBase: A multiuser hypertext database. *SIGMOD RECORD*, 22 (4), 19-25.

[Young 1990] Young, D. (1990). *The X Window System: Programming and Applications with Xt. OSF/Motif Edition*. Englewood Cliffs, NJ: Prentice-Hall.

Acknowledgements

This research was supported in part by the Texas Advanced Research Program under Grant No. 999903-230.

The Bumpy Road of Electronic Commerce

Andrew Odlyzko
AT&T Labs - Research
amo@research.att.com

Abstract: Electronic commerce is widely expected to promote ``friction-free" capitalism, with consumers sending software agents to scour the Net for the best deals. Many distribution chains will indeed be simplified and costs substantially reduced. However, we are also likely to see the creation of artificial barriers in electronic commerce, designed by sellers to extract more value from consumers. Frequent flyer mileage plans and the bundling of software into suites are just two examples of the marketing schemes that are likely to proliferate. It appears that there will be much less a la carte selling of individual items than is commonly expected, and more subscription plans. Therefore many current development plans should be redirected. Electronic commerce is likely to be even more exasperating to consumers than current airline pricing, and will be even further removed from the common conception of a ``just price." As a result, there are likely to be more attempts to introduce government regulation into electronic commerce.

Footnote: This paper incorporates material from an earlier article on electronic publishing, [Odlyzko].

1. Introduction

Electronic commerce (or ecommerce for short) is still small, at least if we consider only online consumer transactions, such as ordering a book from amazon.com over the Internet. In a broader sense, ecommerce is much larger, since financial, news, and legal information services such as Bloomberg, Reuters, and Lexis have total revenues in the billions of dollars. In a still broader sense, electronic funds transfers are already huge, with daily transactions in the trillions of dollars. All these types of transactions are expected to grow, and to become part of a much larger and uniform system of electronic transactions. (For a survey of the current state of ecommerce, and expectations for growth, see [Cohen et al. 1996].)

While we are rapidly moving towards the Information Age, food, shelter, and clothing will remain our most important needs. However, their shares of the economy are decreasing, and the information content of their goods is increasing. This is an old trend. Agriculture has moved from being the largest segment of the economy a century and a half ago to a relatively minor industry, dwarfed by the medical sector, for example. Furthermore, the cost of the basic ingredients in cereals and other foods is a small portion of the total price. As a further example of the decreasing value of raw materials and factory labor, a single celebrity is often paid as much for endorsing an athletic shoe model as all the workers in the undeveloped countries who assemble those shoes. We can expect a continuation of this trend, with the work of the ``symbolic analysts" (who, in Robert Reich's terminology, include lawyers, software writers, and advertising executives) increasing its share of the economy.

The main concern of this essay is electronic trade in information goods, such as news, novels, software, music, movies, as well as legal, medical, and credit information. How will these goods be distributed, and how will their production be financed? Esther Dyson [Dyson 1994] predicts that almost all intellectual content will be available for free. In her view, some content production will be supported by outside advertisers (who already pay for most of the cost of newspapers, for example, as well as all the costs of the commercial TV networks). Some content will likely be made available for free, as a form of advertising for other services by the producers (as the Grateful Dead do in encouraging people to tape their performances, in the hope this will bring more people to their concerts). While Dyson's vision will come true for a large part of the material on the Net, it seems unlikely that it will be universal. Movie studios such as Disney attract large paying audiences to theaters and purchasers to their videotapes through the quality of their products, and are likely to do so in the future. While some novelists make more money from selling movie rights to their plots to Hollywood than from royalties on books, this is rare. Each year, over a hundred times as many books are published as there are movies produced, and books bring in much more money than movie theater tickets. Thus we can expect that content producers will usually want to be paid directly for their work, as that will be the only feasible route to earning a living. Furthermore, Dyson herself [Dyson 1994] emphasizes that much of the value on the Net ``will go to the middlemen and trusted intermediaries who add value - everything from guarantees of authenticity to software support, selection, filtering, interpretation, and analysis." How will these middlemen be paid? It seems likely that often they will wish to collect payment directly from consumers, just as the online legal information service Westlaw collects fees from attorneys who use it. The basic data in Westlaw is court opinions, which are freely avaialable. What gives Westlaw its lock on the market is the control of its citation system.

Many of Dyson's predictions are likely to come true. In particular, huge amounts of intellectual property will be available for free. I expect that this will apply to most scholarly publications, since their authors typically do not receive direct financial benefits from their papers, and are interested in maximizing the circulation of their results [Odlyzko 1996]. However, it seems likely that there will also be a flourishing ecommerce sector, with individuals purchasing goods and services. The question is, how will ecommerce be conducted?

The usual expectation is that ecommerce will promote ``friction-free capitalism," (cf. [Gates 1995]), with distribution costs reduced. It is easy to see how this can happen, as the older communication systems such as the post office, the telegraph, the telephone, and the fax have all served to make the economy more efficient. The Internet creates many more possibilities for improving life. Classified ads, for example, bring in a large fraction of the revenues of the newspaper industry, but can be replaced by a much cheaper and more convenient electronic system. Other parts of the common vision of ecommerce are more questionable, however, and that is what the rest of this essay will discuss. It is often thought that instead of buying an entire newspaper, readers will pay for those individual stories they are interested in. Someone wishing to purchase a VCR might send an ``intelligent agent" into the Internet to collect bids from suppliers for a unit that meets desired specifications, and then select the best choice. While such scenarios will be feasible technically, it is extremely unlikely they will be dominant. Instead, we are likely to see a proliferation of policies such as those of current music CD retailers who sell on the Internet. Most of them do not allow software agents to collect their prices. We are also likely to see a strengthening of the trend towards subscription services and bundling of products, as is done in software suites today. This will often require redirection of development efforts.

This essay is devoted largely to an explanation of the economic reasons that are likely to lead to the creation of ``bumps" on the electronic superhighway. These reasons operate already in the current economy, and are responsible, for example, for the U.S. airline pricing system, which is a source of frequent frustration and complaints. In ecommerce, frustration and complaints are likely to be even more frequent. The reasons for this are twofold. On one hand, the economic incentives to create artificial barriers will be greater in ecommerce than today, since essentially all costs will be the ``first-copy" costs of creating goods, and distribution will be practically free. On the other hand, it will be much more transparent that the barriers are artificial. This will often collide with popular notions of what is fair, and is likely to lead to attempts at much more intrusive government regulations than we have seen so far. In the past governments have been involved primarily in security issues of the Net, and more recently have gotten concerned about pornography. However, in the future they are likely to attempt to regulate the conduct of business on the Net as well.

If the predictions of this essay come true, then some of the current development efforts will turn out to be misdirected. Many systems are planned under the assumption that ecommerce will operate through ubiquitous micropayment schemes, with information goods sold in small units at extremely low prices. Certainly some products will be sold this way, but the arguments in this essay show that much information will be sold via subscription and other more complex marketing mechanisms. This will require different business cases and distribution mechanisms. These arguments also suggest that it will be necessary to prepare to comply with edicts from various governments, edicts that will be be changing and will often be inconsistent.

2. Natural and Artificial Barriers in Commerce

Capitalism is excellent at inducing people to reduce barriers to commercial activities. However, it also produces incentives to create artificial barriers. Some of the barriers are created by government action, such as those of patent and copyright laws, which give owners of intellectual property a limited legal monopoly on the uses of their creations. Other barriers are created by merchants. It is common for an airline passenger to have paid 5 times as much as the person in an adjacent seat, with the only difference between the two being that the first one is not away from home on a Saturday night. The airlines would like to charge the business travelers (who are presumed to be able and willing to pay) more than vacationers (who might drive a car instead or not travel at all), but do not have a direct way to do so. Therefore they impose the Saturday night stopover restrictions to distinguish between those two classes of customers. There have been several attempts by airlines to move towards a simpler system of uniform pricing (sometimes by newcomers, such as People Express, sometimes by established carriers), but they all collapsed. This suggests that there is an underlying economic logic behind this system, however exasperating the results might be. If that is correct, though, we can expect similar moves in ecommerce.

The general tendency in the marketplace is to avoid ``commoditization," in which there are many almost equivalent products and services, and where price is the only consideration. Ford does not compete with Honda in producing the most inexpensive Accord. Instead, it offers the Taurus as an alternative, and there are many features in which the Accord and Taurus differ. Sometimes commoditization is hard to resist. In some cases this happens because consumers learn there is little to differentiate products. Oil companies have pretty much given up on trying to convince people that gasoline differs in anything other than octane ratings. In other cases, commoditization is forced on an industry by government edict or effective private monopoly. Intel and Microsoft have reduced the IBM-compatible PC industry to a commodity business, in which they collect almost all the profits, and the other players scramble to find a niche that will enable them to do more than just break even. However, those are the exceptions. The general ecological principle is towards evolution of species that fill different roles. Zebras do not attempt to compete with giraffes, but exploit a different part of the ecosystem, and evolution does not lead to a convergence of those two species. Similarly, in the world of business, companies try to differentiate their products. Workstation producers could never in the past agree on a common version of Unix, even under the threat of being overwhelmed by PCs, since that would have required giving up the distinctive features that bound them to their customers. Even airlines, which are basically in the commodity business of moving people from one city to another, try to differentiate themselves through frequent flier plans and special pricing schemes.

Ecommerce is likely to lead to a proliferation of pricing plans that will seem to most people to be much more frustrating and less rational than even today's U.S. airlines. There will probably be a niche market for people who care most about their convenience, and will use their intelligent agents to do their shopping for them. However, what Sony, for example, might do is sell to that market only models of VCRs that are not available elsewhere, and are hard to compare to those sold in other places. Stores that have physical buildings are likely to serve a different clientele, and might also take further steps to differentiate themselves to prevent comparison shopping, which will be much easier with many people sharing their experiences on the Internet. There is likely to be a proliferation of frequent-shopper plans. Further, Sony VCRs sold in Sears stores might be slightly different from those sold in WalMart, and model numbers and features might change rapidly to inhibit consumer rating services (such as Consumer Reports, or various Internet-based group-rating schemes that are beginning to develop). There are already artificial barriers to free information flow. Grocery stores routinely bar employees of other stores from collecting extensive data on prices. The policy of Internet CD stores of preventing software agents from collecting prices for comparison shopping is just an extension of such barriers to free information flow to ecommerce. We can expect more such barriers.

While barriers to commerce of the type discussed above are usually perceived as unfair (an issue that I will deal with more extensively in the last section), they can increase not just the producers' wealth, but economic efficiency and social welfare. As a simple example, consider an independent consultant who can produce a technical report that two different customers might be willing to pay $3,000, and $2,000 for, respectively. If she has to charge a uniform price to the two customers, the most she can get is $4,000, obtained by pricing the report at $2,000. However, if she charges the first customer $3,000, and the other $2,000, she will earn $5,000. If the consultant's time and expenses to prepare the report are worth $4,500, she will not undertake the effort if a uniform price is required. From an economic viewpoint it is therefore advantageous to allow her to charge different prices to different customers. However, the customer who pays $3,000 is likely to resent it if somebody else obtains the same product for $2,000, and often will not agree to the deal if all conditions are publicly known. This is caused by a conflict between notions of economic efficiency and fairness.

There are many examples in the marketplace of behavior that appears even less fair. For example, in 1990, IBM introduced the LaserPrinter E, a lower cost version of its LaserPrinter. The two version were identical, except that the E version printed 5 pages per minute instead of 10 for the regular one. This was achieved (as was found by independent testers, and was not advertised by IBM) through the addition of additional chips to the E version that did nothing but slow down processing. Thus the E model cost more to produce, sold for less, and was less useful. However, as Deneckere and McAfee show in their paper [Deneckere & McAfee 1996], which contains many more examples of this type (referred to as ``damaged goods"), it can be better for all classes of consumers to allow such behavior, however offensive it might be to the general notions of fairness. Consumers who do not need to print much, and are not willing to pay for the more expensive version, do obtain a laser printer. Consumers who do need high capacity obtain a lower price than they might otherwise have to pay, since the manufacturer's fixed costs are spread over more units.

Barriers in commerce are an essential part of the current marketplace. Consider the book trade. Although people do not think of it this way, current practices involve charging different prices to different users, and thus maximizing revenues. A novel is typically published in hard cover first, with the aim of extracting high prices from those willing to pay more to read it right away. Once that market is fully exploited, a somewhat cheaper trade paperback edition is made available, to collect revenue from those not willing to pay for the hardbound copy. Finally, a regular paperback edition is published at an even lower price. The used book market develops in parallel, for those willing to read books marked up by previous owners, and so on.

How will ecommerce affect book publishing? Eventually we can expect that all books will be available electronically (and will evolve towards new forms, made possible by digital communications). Costs of publishing will come down, and this is going to increase the supply, and lead to many works distributed for free, by aspiring authors hungry for the recognition that might lead to fortune. What about those electronic books that people will be willing to pay for? With publishing costs reduced, we can expect that the authors' share of the revenues will rise, say from the current 15% or so royalty rate to 50% or more, and so in effect the authors might become much more influential than the publishers (or might become the publishers themselves). However, since publishers obviously benefit from the present system of differential pricing, they (and the authors) are likely to have an incentive to institute a similar system in the digital arena. The issue is how to do this. Bits are bits, after all, and are easy to copy.

If we make only simple extensions of current copyright laws, we are likely to see a great change in the marketplace for information goods. When I buy a book, I cannot make a copy of it and sell that copy to somebody else. On the other hand, I can sell, rent, or give away the book I purchased to anyone I wish. Suppose we carry over exactly the same rights to the digital world, with some combination of cryptographic techniques and laws guaranteeing that unauthorized copies of digital ``books" cannot be made. The ease of transactions on the Net (which is what leads to the dreams of ``frictionless capitalism") would then force major changes. With physical volumes, there are substantial barriers to trade in books. Most people do not like reading books that are tattered or marked up by others. They take their time reading books, and (especially for the ones they enjoy) like to retain them in their libraries to be reread any time they wish. As a result of these natural barriers, a single copy is usually read by only a few people. The economics of the present book publishing business depend on this phenomenon. In the digital world, though, with high bandwidth networks and efficient intermediaries, I could buy a copy of a book an hour before bedtime, read a new chapter, and then, just before turning off the lights, send that copy off for resale. Instead of a million copies of a printed book, a thousand electronic copies might suffice. This would force a dramatic change in the structure of the book publishing industry, and explains why there is an intense interest in the creation of artificial barriers to ecommerce, either through revisions to copyright laws or through technological methods.

3. The Bumps on the Electronic Highway

Some types of barriers to commerce are accepted as natural when dealing with physical goods. It would be prohibitively expensive for the New York Times, say, to distribute 100 little sheets each day, each one with a separate story, and having readers buy just the ones they were interested in. The accepted wisdom is that ecommerce will lead to the electronic equivalent of just that, with readers selecting and paying for individual stories. It will certainly be possible to do so, as micropayment systems are being developed that will allow for processing of tiny transactions, such as payment for a single story in the New York Times, or a ``hit" on some aspiring poet's Web page. However, the economic argument is that while such schemes might exist, and may be used in some situations, they will not be dominant. The example of book publishing in the previous section shows why producers of information goods benefit from the natural barriers that exist in the physical world. Their incentives to create artificial barriers in the digital world will be even stronger. It will be harder to distinguish between consumers, since transactions will tend to be impersonal, and arbitrage will be easy. Most important, distribution costs will be negligible, so that only the ``first copy" cost of creating a work will matter. Hence traditional, commodity-market type of competition, in which the market price equals the marginal cost, will have to be avoided, since marginal prices will be essentially zero. The incentive that low marginal costs provide to create barriers in commerce can already be seen in many high technology fields. The ``damaged goods" studied in [Deneckere & McAfee 1996] come primarily from such areas. The pharmaceutical industry is notorious for selling products for hundreds of times more than the cost of producing them, and for selling the same chemicals for human use for ten times the price charged for veterinary purposes.

While the incentives to erect artificial barriers will be large in ecommerce, there will also be novel possibilities created by the electronic medium. What kinds of barriers are we likely to encounter in ecommerce? The four most important ones will probably be bundling, differential pricing, subscriptions, and site licensing. That they are likely to be prominent in ecommerce has been pointed out before, especially by Hal Varian [Varian 1995a], [Varian 1995b]. In the rest of this section I will explain how they operate, and why they are attractive to content producers. There are additional arguments in favor of subscription and site licensing plans. For example, security problems are likely to be easier to solve in those cases. It is also much easier to learn about customers' preferences and tailor special offerings for them when they are subscribers as opposed to anonymous purchasers of individual items. However, this essay will deal only with the basic economic arguments.

The basic assumption in the economic analyses below is that for each information good, an individual consumer will purchase it only if the price is below some threshold (that consumer's valuation of the good). For simplicity, I will only consider items that are independent of each other (such as stories in a newspaper). Much of the economic literature cited below is concerned with goods that are related in one way or another. (For example, if I buy a spreadsheet from Corel, I am unlikely to purchase another one from Microsoft. On the other hand, if I buy a presentation package, I am more likely to buy a CD-ROM of pictures than I would otherwise.) I will not take these factors into consideration, to keep the presentation simple, and bring out only the main factors that are likely to influence the development of ecommerce. I will also assume, as is standard, that producers cannot in general find out what an individual is willing to pay for a product, but can, through test marketing, say, obtain an accurate statistical description of the valuations that the whole population of potential buyers place on that product.

3.1. Bundling

Bundling consists of offering several goods together in a single package, such as combining a word processor, a spreadsheet, and a presentation program in a software suite (such as Microsoft Office), or else printing many stories in a single newspaper. Bundling is common, and often seems natural. For example, right shoes and left shoes are invariably sold together, and just about the only time anyone might regret this is when a dog chews up one of a new pair of shoes. I will concentrate on bundling of goods that are almost unrelated, such as a word processor and a spreadsheet program. Why should the pair of them together sell for much less than the sum of their separate prices? It is useful to have seamless integration of the two, to make it easier to move material between them, to have common command structure and icon layouts, and so on. That seems to argue for charging more for the bundle than for the pieces! However, bundling, with a lower price for the bundle than for the components, or even without any possibility for purchasing the components separately, is common. The reason is that it allows the producer to increase revenues by capturing more of the ``consumer surplus" that arises when customers pay less than they are willing to do. Since in general prices have to be the same for all customers, bundling can be used to smooth out the uneven preferences people have for different goods and services. For example, suppose we were dealing with a proposal to start a newspaper that would have two sections, a business page and a sports page. Suppose also that there were just two potential readers, Alice and Bob. Suppose also that Alice needs to keep up with the financial world, and so is willing to pay $0.50 for the business page, but only $0.20 for the sports page, since she does not particularly care about sports, but might like to keep up with lunchtime conversations. Suppose that Bob's preferences are reversed, in that he is an eager sports fan, willing to pay $0.50 for the sports page, but only $0.20 for the business page, since all he cares about is occasionally checking on his retirement fund. Under those conditions, how should the proposed newspaper be priced? If each section is sold separately, then a price of $0.20 for each will induce both Alice and Bob to buy both sections, for total revenues of $0.80. If the price is set at $0.50 for each section, then Alice will buy only the business page, and Bob only the sports page, for total revenue of $1.00. On the other hand, if the two sections are bundled together, then a price for both of $0.70 will induce both Alice and Bob to purchase the newspaper, and will produce total revenues of $1.40. Thus the economically rational step is not to offer the two sections separately, but only bundled together.

Bundling has been studied extensively in the literature, starting with the paper of Burnstein [Burnstein 1960]. A few other references are [Adams & Yelen 1976], [Bowman 1967], [Economides 1993], [Krishna et al. 1996], [Schmalensee 1982], [Stigler 1963], [Varian 1989], [Wilson 1993], [Wilson 1996]. Unfortunately there is no simple prescription that can be given as to when bundling is better than selling items separately. Depending on the distribution of consumer preferences, bundling can be either more or less profitable for the producer, as was already shown by Adams and Yellen [Adams & Yelen 1976]. However, there are some general guidelines. One is that bundling becomes more profitable as marginal costs decrease. (This may partially explain why software suites spread at about the same time as unpaid support provided to users by software houses decreased.) Another is that bundling becomes more attractive when consumer preferences are negatively correlated (as in the example above, where Alice and Bob had almost opposite tastes). However, negative correlation in valuations is not necessary for bundling to be profitable, as was first pointed out by Schmalensee [Schmalensee 1982], and as will be shown in the example below. Random variations in preferences are sufficient as a result of the law of large numbers.

How much of a difference can bundling make to a producer's bottom line? Unfortunately the published literature is practically silent on this point, for reasons I will discuss later. (There is one intriguing computation in [Stigler 1963], based on reported revenues of movie theaters in different cities.) Let us therefore consider some artificial examples, a bit more realistic than the Alice and Bob one presented above. Consider two books, A and B, say ``The Tannu-Tuva Cookbook" and ``Sherlock Holmes in Antarctica." Suppose that among one million potential customers, book A is valued at $1 by 100,000, at $2 by another 100,000, and so on, up to $10 by 100,000, and suppose the same distribution of valuations applies to book B. Suppose further that the valuations of the two books are independent. Thus there are about 10,000 customers who value book A at $3 and simulataneously book B at $5, and similarly about 10,000 customers who place values $9 and $2 on A and B, respectively. Under these conditions, if the publisher is to sell these books separately, revenue will be maximized when the price of each is set at $5. About 600,000 people will purchase each book, for total revenue from sales of both books of $6,000,000. (This maximum is not unique, as the same revenue can be achieved by pricing each book at $6, in which case about 500,000 people will buy each.) However, if the two books are sold together, revenue can be made much higher. Since there are 10,000 people who value the bundle at $2 (exactly the 10,000 who value each book at $1), while there are 90,000 who value it at $10, a short calculation shows that the revenue-maximizing price is $9. At the price of $9 per bundle, 720,000 people will purchase it, for total revenue of $6,480,000, exactly 8% higher than if the books were sold separately. Since profits are the revenues minus the fixed costs of producing the books, they would increase much more dramatically.

What weakens the case for bundling is that most people have no interest in most goods. In the example of the books ``Sherlock Holmes in Antarctica" and ``The Tannu-Tuva Cookbook," a more realistic assessment would be that in a population of 1,000,000, each book would be valued at zero by 90% of the population, with 10,000 valuing it at $1, 10,000 at $2, and so on. If the 100,000 people who do place a positive value on book A are distributed independently of those who value book B at $1 or more, then there are only 10,000 people who place positive values on both A and B. Bundling under these conditions does not produce much benefit. However, even in cases of extreme indifference, bundling can be profitable if there are enough goods. Consider an information service with 1,000 items (news stories, pictures, or songs). Suppose that in a large population, each individual is totally uninterested in 900 of the items, and values 10 at $0.01 each, 10 at $0.02 each, and so on, with 10 valued at $0.10 each. If the items are to be sold individually, a revenue-maximizing policy is to charge $0.05 for each. Each customer will then purchase 60 items for a total of $3.00. However, if the collection is sold as a whole (which involves no extra cost to producers in case of information goods, and also no cost of tossing out mounds of unwanted boxes to consumers), then a price of $5.50 will induce each person to buy, for a gain of 83% in revenues (and more in profits).

So far we have compared only sales of unbundled products (pure unbundling) to those of bundles (pure bundling). However, it is often advantageous to use mixed bundling, where both bundles and separate goods are offered. In the example of the books ``Sherlock Holmes in Antarctica" and ``The Tannu-Tuva Cookbook," with the distribution of valuations assumed above, a price of $10 for the bundle and $5 for each book separately would produce revenue of $7,400,000, about 14% higher than pure bundling, and over 23% higher than pricing the books separately. (Note that the optimal combination above has the paradoxical property that the price of the bundle is exactly the price of the pieces. Under the assumption of the model, people who value book A at $7 and book B at $3 will purchase the bundle, but if the bundle is not available, will only purchase A.) Adams and Yellen [Adams & Yelen 1976] have shown that mixed bundling is always more advantageous to the producer than pure bundling.

Toy models like the one above are amusing to play with, and help illustrate the advantages to producers of bundling. If the distribution of consumer valuations is known, one can determine numerically what the optimal policy is for the producer [Wilson 1993], [Wilson 1996]. Unfortunately the basic assumption that consumers know what value they place on various goods, and purchase them precisely when the price is below their value, is questionable. In practice people behave in much more complicated ways. An old joke illustrates this:

> Waiter: And for dessert, we have chocolate mousse, apple pie, and ice cream.
>
> Customer: I will have apple pie.
>
> Waiter: Oh, I forgot to mention that we also have Peach Melba.
>
> Customer: In that case I will have the mousse.

While this is a joke, actual behavior is often just as paradoxical. Catalog merchants have learned that the attractiveness of an item is affected strongly not just by its price and description, but also by its placement among other offers. Consumer choices are complicated. Some of the seemingly irrational behavior can be explained on the basis of different consumers having different sensitivities to prices. For example, the phenomenon of regular sales has been modeled successfully this way in [Varian 1980] and later papers. Other interesting phenomena emerge if one assumes that consumers do respond to price signals in an economically rational way, but with some delay (see [Richardson & Radner 1996], for example). However, there is no complete theory. Experimental economics has shown that economically optimal solutions can be attained even with small groups of agents, provided they are working in a constrained environment and are trying to optimize their wealth, although even there paradoxes abound (cf. [Cook & Levi 1990], [Hagel & Roth 1995]). In general settings, though, human behavior is hard to model. There are nontransitivities in preferences, choices are determined by behavior of others (so a person is more likely to see a movie that colleagues have seen to have something to talk to them about as well as because that person is likely to trust their judgement), and so on. Companies collect extensive data from test marketing, but that data is noisy, and typically involves only small variations in test parameters. There seems to be no unambiguous empirical demonstration that a well defined demand curve exists. Thus economic models discussed above do indicate that bundling is likely to be advantageous to producers, but do not prove this.

What happens in the real marketplace, with a variety of customers and competitors, and where there is already much experience with a variety of marketing plans? What we see there is extensive evidence of bundling, which confirms the prediction of the economic models. In many situations, such as that of physical newspapers, there is an obvious motivation for bundling to reduce costs. However, there is also evidence of bundling's success when there are practically no physical costs involved. Software suites such as Microsoft Office are just one example. Cable TV does not charge for each channel separately, but for packages (bundles) of them. Finally, the big and profitable online information services in the financial and legal arena, such as Reuters, Bloomberg, and Lexis, all operate on a subscription basis or appear to be moving in that direction. (The ``pay-per-view" approach made more sense when the computing infrastructure for online access was expensive, and therefore there were high marginal costs of providing access.) All this evidence confirms that bundling is likely to be common in ecommerce.

3.2. Differential Pricing

Charging different prices to different consumers is already common. Various senior citizen or student discount programs are just some of the most widespread practices. Scholarly journals typically charge much higher prices to libraries than to individuals, sometimes 10 times higher. For a thorough discussion of such price discrimination and its economic and legal status, see the survey [Varian 1989]. A producer would like to charge according to the consumer's willingness to pay, but the consumer will usually be reluctant to reveal such information.

However, it is sometimes possible to correlate willingness to pay with other features. Airlines offer much cheaper tickets for those willing to be away from home on Saturday night. The theory is that business travelers, who are willing to pay a lot, will not be willing to put up with such inconvenience. In information services, online services such as Prodigy and CompuServe offer stock market quotes that are delayed by 15 or 20 minutes for no extra cost, beyond the basic subscription. Real-time quotes uniformly cost extra, on the theory that those who need them for their trading will pay more.

The software industry relies on differential pricing in many products. Student or demo versions typically are the same as the main packages, except for artificial limitations on what they can do. They either cannot produce large executables, or cannot handle large files, or cannot use extended precision. We are likely to see many more examples of such differential pricing. Electronic publications may offer a high-resolution version at one price, a lower-resolution version at a lower one, and sometimes might offer a fax-quality version at no charge. There are already interesting experiments in book distribution, with authors making some parts of their manuscripts freely available on the Internet, to advertise their work, to update it with lists of current errata, and to make available features that draw on the unique capabilities of the electronic medium. There are also likely to be differentials based on timeliness, as with stock market quotes; old issues might be offered at low or no charge. There might be extra charges for links to cited works or other desirable features.

Differences in quality of offered products might be the only way to preserve some of the features of public libraries. In the digital realm, without some artificial barriers, there would be practically no difference between buying and borrowing. Hence the traditional library policy of unrestricted lending is not compatible with ecommerce, and we are likely to see artificial barriers. Databases might be available to library customers but only inside the library, at special terminals, for example. Librarians would then become gatekeepers, restricting access to material more than making it freely available.

3.3. Subscription vs. Pay-Per-View

Offering access to a database or a movie channel on a subscription basis is a form a bundling. The alternative is to charge for each movie, or each download of a Web page. There is much discussion of how such ``a la carte" shopping might become prevalent. One attraction of programs consisting of small applets that can be downloaded on demand appears to be the perception that this would allow producers to charge according to how frequently the software is used. However, past experience with pay-per-view systems has been discouraging. Except for a few events, such as championship boxing matches, they have not succeeded in attracting much revenue. All the arguments in favor of bundling apply, and suggest that pay-per-view systems will not be common. Furthermore, there are additional arguments, supported by empirical data on consumer behavior, that argue against pay-per-view schemes. Consumers appear to have a strong predilection for reducing risk, even when this predilection results in lower than optimal expected financial payoff. A certain $10 gain is usually preferred to a wager in which there is a 90% chance of winning $15 but also a 10% chance of losing $20, even though the latter has expected payoff of $11.50. People also tend to use small deductibles when purchasing fire or casualty insurance, even when they could easily bear the loss from a larger deductible. (Since few insurance companies operate with an overhead of less than 30%, a larger deductible would almost surely lead to savings in the long run.)

Similarly, consumers appear to have a strong preference for subscription services. To a large extent this is probably explainable by general risk aversion. I may prefer to pay a higher price for a word processor now, even if I do not need it much, to have free use of it when I lose my job, and need to send out lots of job applications, but will not be able to afford extra charges. This preference for subscription services is present even among librarians, who are not spending their own money, and with a large number of users of their resources might be expected to have a stable and predictable usage pattern. Even so, they have often expressed their unease about paying ``a la carte" for access to databases, since they feared they could not predict what this would do to their budgets. It is difficult to quantify the strength of this preference for subscription services, but it exists and is strong. In the 1970s, the Bell System experimented with charging for local calls. Typically, customers were given a choice of the traditional flat rate option, which might cost $7.50 per month, and allow unlimited local calling, and of a measured rate option, which might cost $5.00 per month, allow for 50 calls at no extra charge, and then cost $0.05 per call. Anyone making fewer than 100 local calls per month was better off with the measured rate option. Careful studies of consumer behavior were carried out by Bill Infosino, Gerry Ramage, John Rotondo, and others at AT\&T. They observed that typically around 50% of the customers who were making almost no local calls at all, and thus would have benefited from measured rate service, still stayed with the more expensive flat rate service. The preference for flat rate pricing for Internet access is another example of this phenomenon.

The main conclusion to be drawn from this discussion is that subscription services do offer substantial value to consumers, even if that value may seem to be irrational. As a corollary, they also offer value to producers. People are willing to pay a lot just to be able to occasionally use certain features. Software producers complain about all the heavy users of their products who do not pay for their high usage. However, these producers benefit from the many users who hardly ever use their system. I seldom use Microsoft Word, but when I do use it (typically because somebody sends me a Word document), I do need it, and so am willing to purchase it for just such occasions. Hence we can expect that even if large systems consisting of downloadable applets do become practical, they will be available on a subscription, and not on a per-use basis.

3.4. Site Licensing

Site licensing, in which a company or a university pays a flat fee to allow everyone in that institution to use some program or access a database, is very common in the computer and online information industries. In some forms, it has been present for a long time in other areas as well. For example, scholarly publishing can be thought of as an example of site licensing. Typically a university will buy a single copy of an esoteric journal, which is then placed in a library, to be consulted by anyone on campus.

In software, site licensing has many attractive features. It simplifies the enforcement problem (which is nontrivial, since many corporations report they spend more on policing software use than on the purchase of that software). It also encourages new users to try out a package, and thus stimulates more usage. In addition, though, site licensing has a strong direct economic argument behind it. We can think of site licensing as a variant of bundling. In ordinary bundling, a producer assembles together several goods into a bundle, to smooth out the differences in valuations that individual consumers place on those goods. In site licensing, a producer assembles together a group of consumers to smooth out the differences in valuations that different people place on a single product. As an example, suppose that in a company of 1,000 employees, 900 are totally uninterested in a software package, but 10 feel it is worth paying $10 for it, 10 feel it is worth $20, and so on, up to 10 who feel it is worth $100. If the software manufacturer had to sell copies of the package to individuals, the best price would be either $50 or $60 for a copy, and the revenue in either case would be $3,000. However, if the management of the company has an accurate estimate of how much the employees value the product, it should be willing to pay $5,500 for a site license. This would be a much better deal for the producer, even though it would bring in only $5.50 for each person entitled to use the product. Hence we can expect further spread of site licensing. (For some other aspects of site licensing, see [Varian 1995b].)

4. Fairness, Legality, and Efficiency

Economic arguments show that there is value to many of the artificial barriers in commerce. It is value not just to producers of the goods and services, but to society. Moreover, the incentives to create such barriers apply to individuals as well as large corporations. If Alice plays the piano, and Bob performs magic tricks, they might be able to obtain a higher total income by bundling their services through offering a combined act to nightclubs. The result might be the difference between starvation and relative comfort. In ecommerce, a group of budding poets might collect larger revenues if they sell access to their combined works, instead of acting individually.

While economics will lead to the creation of barriers in ecommerce, this will frequently clash with popular notions of what is fair. There is already much grumbling about airline pricing and senior citizen discounts. Moreover, many of the grumbles result in laws restricting commerce. Several cities in the United States have passed laws decreeing that women's shirts should not cost more to launder than men's. There is a general perception of what is fair, often codified into laws. Some is based on ideas of non-discriminatory treatment (as with laundry pricing practices). Some goes back to the ancient notion of a ``just price,'' which is supposed to reflect a modest markup over the producer's costs. However, in ecommerce, even more than in the modern physical economy, cost is a poorly defined concept.

In ecommerce, the concepts of ``increasing returns'' [Arthur 1994], in which producer profits increase as usage increases, and customer lock-in, in which someone trained in using a particular spreadsheet faces a major barrier of retraining in switching to another one, are among the ruling ones. This means that the many traditional tests of illegal monopolistic behavior do not apply. It can make excellent sense to give away a software package, since the major benefit to the producer will come from sales of upgrades. Other examples of economically sensible behavior that is not accepted by society exist. U. S. courts stopped IBM from requiring users of its tabulating machines to purchase their punched cards from IBM [US 1936]. Today, most economists would argue that this decision was a mistake, since in effect what IBM was attempting to do was to charge the heavy users more than the light ones, to enlarge the market. (See [Stigler 1963] for economic arguments against another decision, [US 1962], which barred movie distributors from requiring movie theaters to book whole series of movies instead of selecting them individually.) While the general issue of what practices are legal is at best murky (cf. [Bork 1993], [Bowman 1967], [Varian 1989]), there may be legal problems with some of the barriers that are likely to be erected. Even when there is no legal difficulty, there can be extensive public action, as in recent protests against pharmaceutical firms' pricing, and against use of child labor in less developed countries. (With reputations, whether of celebrity endorsers or producers themselves, becoming increasingly important, public protests can be powerful weapons.) Issues of fairness (see [Zajac 1995] for extensive discussions of their influence on public policy) are likely to be much more pronounced than in the past. One reason is that the barriers on the electronic superhighway are likely to be frequent. Another is that those barriers will be much more visible as artificial. In print book publishing, most people seem to think that hardcover books sell for more than paperbacks because they cost more to produce. However, the differences in costs are minor, and the price difference is just a form of price discrimination. On the Web, it will be clear that a low resolution version of a work is just a degraded version of the high resolution one. It will also be much easier to organize protest movements than in the past.

Public perceptions of what is fair depend on culture, are often inconsistent, and do often clash with economic incentives. Furthermore, the rapid evolution of technology, markets, and laws, will lead to a continuation of the unstable situation we have. There may be serious protests against the ``winner-take-all'' society [Frank & Cook 1995] that electronic commerce might be seen to promote, where millions of aspiring novelists work hard to catch the public's attention, but a small handful manage to catch all the material rewards. Even without general protests, there will be increasing temptation to ask governments to intervene, and that will produce serious difficulties for ecommerce. Barlow's ``independence declaration'' [Barlow 1996] might appeal to many, but is totally unrealistic. Government has been involved in setting up the Internet, and is getting more involved all the time, through issues such as the fair use of Scientology documents on the Net, assignments of names, and provision of wide access to the Net. The U. S. Telecommunications Act of 1996, which nominally deregulated telecommunications, also brought in extremely intrusive government regulations, to deal with thorny issues of setting up a ``level playing field." We should be prepared for more intervention of this type, whether they are successful or not.

Many issues will be complex. As an example, only a tiny fraction of the public understood any of the arguments about the U. S. telecommunications deregulation debate, with its technical points about access to local wires. Also, few people follow the details of the debate about revisions to copyright laws. As was argued in an earlier section, ecommerce requires some revision. However, there are a variety of ways to do this, and the precise ways in which different proposals affect different players is not clear to the public. (See the discussions by Samuelson [Samuelson 1996a], [Samuelson 1996b] of the proposed revisions to U. S. copyright law [USPTO 1995], as well as the survey paper [Okerson 1996] and the book [Patterson & Lindberg 1991].) Therefore we can expect an increased demand for lobbyists, lawyers, and public relations experts. Even in the non-governmental arena, it is reported, for example, that ``in preparing a commemorative CD-ROM for the 500th anniversary of the first Columbus voyage to America, IBM spent over $1M clearing rights, of which only about $10K went to the rights holders; everything else went into administrative and legal fees" [Lesk 1995]. Although systems are being developed for automatic tracking of rights to copyrighted material and the automatic payment of fees, it is unlikely that such systems will see wide usage. Content owners will probably be reluctant to rely on them, and possibly let valuable rights slip away.

The conclusion to be drawn from this essay is that electronic commerce will increase the efficiency of the economy. However, it will also create artificial barriers, and we will have to learn to live with them.

References:

[Adams & Yellen 1976] Adams, W. J., & Yellen, J. L. (1976). Commodity bundling and the burden of monopoly, Quart. J. Economics, 90, 475-498.

[Arthur 1994] Arthur, W. B. (1994). Increasing Returns and Path Dependence in the Economy, U. Michigan Press.

[Barlow 1996] Barlow, J. P. (1996). A cyberspace independence declaration, email broadcast message of Feb. 9, 1996, available at URL http://syninfo.com/IAN/02136002.htm and many other Net sites.

[Bork 1993] Bork, R. (1993). The Antitrust Paradox, 2nd ed., Free Press.

[Bowman 1967] Bowman, W. S., Jr. (Nov. 1967). Tying arrangements and the leverage problem, Yale Law J., 67, 19-36.

[Burnstein 1960] Burnstein, M. L. (1960). The economics of tie-in sales, Rev. Economics and Statistics, 42, 68-73.

[Cook & Levi 1990] Cook, K. S., & Levi, M., eds., (1990). The Limits of Rationality, Univ. Chicago Press.

[Deneckere & McAfee 1996] Deneckere, R. J., & McAfee, R. P. Damaged goods, J. Economics and Management Strategy, to appear.

[Dyson 1994] Dyson, E. (1994). Intellectual value, first published in Dec. 1994 in Release 1.0, republished (in an abbreviated form) in Wired, July 1995, and available at URL http://www.hotwired.com/wired/3.07/features/dyson.html

[Economides 1993] Economides, N. (1993). Mixed bundling in duopoly, working paper, available at URL http://edgar.stern.nyu.edu/networks/cvnoref.html

[Economides 1996] Economides, N. The economics of networks, Intern. J. Industrial Organization, to appear. Available at URL http://edgar.stern.nyu.edu/networks/cvnoref.html

[Frank & Cook 1995] Frank, R. H., & Cook, P. J. (1995). The Winner-Take-All Society, Free Press.

[Gates 1995] Gates, B., Myhrvold, N., & Rinearson, P. (1995). The Road Ahead, Viking.

[Hagel & Roth 1995] Hagel, J. H., & Roth, A. E., eds. (1995). The Handbook of Experimental Economics, Princeton Univ. Press.

[Krishna et al. 1996] Krishna, A., Kopalle, P. K., & Assuncao, J. L. Bundling of complementary goods: The impact of competition, brand preference, and price sensititivy, in preparation.

[Lesk 1995] Lesk, M. (1995). The seven ages of information retrieval, to be published.

[Odlyzko 1996] Odlyzko, A. M. (June 1996). On the road to electronic publishing, Euromath Bulletin, 2(1), to appear, and to be available at URL http://www.math.ethz.ch/~shared/emb Available electronically at URL http://netlib.att.com/netlib/att/math/odlyzko/index.html and via email by sending the message send epublishing.road.txt from att/math/odlyzko to netlib@research.att.com.

[Okerson 1996] Okerson, A. S. (July 1996). Who owns digital works?, Scientific American, 275, 64-68. Text available electronically at URL http://www.sciam.com/WEB/0796issue/0796okerson.html

[Patterson & Lindberg 1991] Patterson, L. R., & Lindberg, S. W. (1991). The Nature of Copyright: A Law of Users' Rights, Univ. Georgia Press.

[Richardson & Radner 1996] Richardson, T. J., & Radner, R. Monopolists and viscous demand, to be published.

[Samuelson 1996a] Samuelson, P. (1996). Intellectual property rights and the global information economy, Comm. ACM 39, 23-28.

[Samuelson 1996b] Samuelson, P. Technological protection for copyrighted works, Emory Law J., to appear.

[Schmalensee 1982] Schmalensee, R. (1982). Pricing of product bundles, J. Business, 57, S211-S230. Comments on S231-S246.

[Stigler 1963] Stigler, G. (1963). United States v. Loew's Inc.: A note on block booking, Supreme Court Review, 152, 152-157.

[US 1936] International Business Machines Corp. v. United States, 298 U.S. 131 (1936).

[US 1962] Loew's Inc. v. United States, 371 U.S. 38, 52 (1962).

[USPTO 1995] Intellectual Property and the National Information Infrastructure, The Report of the Working Group on Intellectual Property Rights, B. A. Lehman, Chair, U. S. Patent and Trademark Office, Sept. 1995. Available at URL http://www.uspto.gov/web/ipnii/

[Varian 1980] Varian, H. R. (1980). A model of sales, Am. Economic Review, 70, 651-659. Erratum on p. 517 of vol. 71 (1981).

[Varian 1989] Varian, H. R. (1989). Price discrimination, 597-654 in Handbook of Industrial Organization, vol. I, R. Schmalensee and R. D. Willing, eds., Elsevier.

[Varian 1995a] Varian, H. R. Pricing information goods, available at URL http://www.sims.berkeley.edu/~hal/people/hal/papers.html

[Varian 1995b] Varian, H. R. Buying, renting and sharing information goods, available at URL http://www.sims.berkeley.edu/~hal/people/hal/papers.html

[Wilson 1993] Wilson, R. (1993). Nonlinear Pricing, Oxford Univ. Press.

[Wilson 1996] Wilson, R. (1996). Nonlinear pricing and mechanism design, 249-289 in Handbook of Computational Economics, vol. I, H. M. Amman, D. A. Kendrick, and J. Rust, eds., Elsevier.

[Zajac 1995] Zajac, E. E. (1995). Political Economy of Fairness, MIT Press.

Acknowledgements: I thank Peter Fishburn, Stevan Harnard, Hsueh-Ling Huynh, Bill Infosino, Steve Lanning, Peter Linhart, Gerry Ramage, Ryan Siders, Hal Varian, and Ed Zajac for their comments and the information they provided.

An Ongoing Experiment in ODL Using New Technologies

Nikolaos Papaspyrou (nickie@softlab.ntua.gr) [1]
Anastasios Koutoumanos (tkout@softlab.ntua.gr) [1]
Hermann Maurer (hmaurer@iicm.tu-graz.ac.at) [2]
Emmanuel Skordalakis (skordala@softlab.ntua.gr) [1]

[1] National Technical University of Athens
Department of Electrical and Computer Engineering
Division of Computer Science
Polytechnioupoli, 15780 Zografou, Greece

[2] Institute of Information Processing and Computer Supported Media
Graz University of Technology
A-8010 Graz, Austria

Abstract: In this paper we present the ongoing project *EONT* - An *E*xperiment in *O*pen and Distance Learning (ODL) using *N*ew *T*echnologies. This project is partly funded by the European Union, within the framework of the SOCRATES programme, and its main objective is to provide an answer to the question whether ODL using new technologies is effective. EONT will achieve its goals through experimentation. We will use an ODL environment based on some of the available new technologies (namely computer networks and networked hypermedia systems), one subject matter area and real learners. Two Distance Learning universities and five conventional universities from seven European countries will take part in this project.

1. Introduction

Contemporary educational systems are criticized as having many drawbacks. Especially conventional universities, as the institutions of tertiary education, have been denounced for the constraints they impose on student selection as well as on the time and place of instruction delivery. *Open and Distance Learning* (ODL), as an instruction delivery method, has been proposed as a means of overcoming constraints and thus contributing in the improvement of education offered by these institutions. As a result, Open universities, Distance Learning universities and Dual-Mode universities have formed and have been in operation for quite a time, alongside with the conventional universities. However, irrespective of their type, the universities' problems have not been resolved up to now, at least not to the degree it was desired.

With the development of new information technologies, especially computer networks and networked hypermedia systems, it seems that time has come for ODL to become more effective. Indeed, the prospects are very good. Various media, like text, sound, still and moving images, etc., can be used in an integrated learning environment alongside with efficient, synchronous and asynchronous communication mechanisms. These technologies can be used in various ways. Although some attempts have started, no results are available as yet proving the effectiveness of ODL using the new technologies. Experimentation is needed to give an answer to this very important question. A project titled "*An experiment in ODL using the new technologies (EONT)*" intends to tackle this problem. This project has been set up and is currently running within the European Union "*SOCRATES Programme*" for Education.

In this paper, an overview of the project EONT is given with particular emphasis on its objectives and method of implementation. No evaluation results are presented since no such results are available as yet. The paper is structured as follows: In [The SOCRATES Programme] we give a brief description of the

SOCRATES programme. In [The Project EONT] we present the project EONT and highlight its tight connection to computer networks and in particular to the Internet. In [Implementation Approach] we describe the approach being adopted for the implementation of EONT. Finally, in [Concluding Remarks], we discuss some issues concerning the project EONT and its expected results.

2. The SOCRATES Programme

SOCRATES [Commission 1995a, Commission 1995b] is the European Community action programme for cooperation in the field of education. Adopted on March 14, 1995 and spanning the period until the end of 1999, it is applicable to the 15 Member States of the European Union as well as to Iceland, Liechtenstein and Norway, in the framework of the European Economic Area agreement. SOCRATES is to be seen in the broader context of promoting lifelong learning in response to the challenge of addressing the constant educational needs resulting from technological change, quickening obsolescence of knowledge, and the role of education in enabling people to fulfill their individual potential. Education and training are of central importance to Europe's economic and social future. In the increasingly competitive and global marketplace, high quality human resources and the effective production, transfer and sharing of information are of paramount importance.

The overall aim of SOCRATES is to help improve the quality and relevance of education for children, young people and adults, by enhancing European cooperation and increasing access to a range of learning opportunities available across the Union. It seeks to provide learners of all ages and social groups with insights into the European dimension of the subjects which they are studying, to increase opportunities for personal experience of other European countries, to develop a stronger sense of sharing a European identity and to foster the ability to shape and adapt to changes in the economic and social environment. Among the eight actions of SOCRATES, one is devoted to ODL. The development of ODL is a key factor enabling citizens of the European Union to create and take advantage of an open area for educational cooperation in Europe. It is also one of the six areas for community action identified by the Maastricht Treaty as being of particular importance for improving the quality of education. The SOCRATES measures, focusing specifically on the support of ODL, are designed to contribute to:

- *Objective A*: facilitating cooperation between organizations and institutions in the field of ODL;
- *Objective B*: enhancing the skills of teachers, trainers and managers in the use of ODL techniques;
- *Objective C*: improving the quality and user-friendliness of ODL products; and
- *Objective D*: encouraging the recognition of qualifications obtained through ODL.

Concerning ODL, the emphasis within SOCRATES is essentially on stimulating European cooperation in:

- the use of distance learning as a means of overcoming barriers to physical mobility; and
- the use of information and communication technology for improving the quality of conventional education.

ODL involves the use of new methods -technical and/or non-technical- to improve the flexibility of learning in terms of space, time, choice of content or teaching resources and/or to improve access to educational systems from a distance. Promoting ODL can refer to:

- improving the quality of the organizational environment in which this form of education takes place; and
- improving the availability and quality of the teaching media and resources for this type of learning.

SOCRATES supports two types of projects in ODL: partnership projects and observatory projects. The purpose of the former, is to reduce fragmentation, to avoid duplication of effort, to improve working methods through exchanging experience and methodologies, and in general to achieve greater synergy at European level by working together on joint transnational projects. The purpose of the latter is to produce a comprehensive picture of the state of development, concerning a particular aspect of ODL or the use of new educational technologies across a broad cross-section of the countries participating in SOCRATES.

3. The Project EONT

EONT (URL: http://hyperg.softlab.ntua.gr/EONT/) is a partnership project between seven universities from seven European Union countries, as depicted in [Appendix A]. Two of these universities are Distance Learning, whereas the rest are conventional. The partnership was formed on the basis of the partners' common interest in experimenting with ODL using the new technologies. The main objective of the project is to *make an experiment* in ODL using the new information and communication technologies, namely computer networks and networked hypermedia systems, in order to explore the effectiveness of such an approach. Other objectives are:

- to exchange information and experience concerning the use of the new information technologies in ODL;
- to stimulate the use of ODL in conventional as well as Distance Learning universities;
- to contribute in the development of the European dimension in education;
- to explore the possibility of using ODL as a means of providing learners in one country with access opportunities to education institutions in another country;
- to identify issues in using the new technologies in ODL; and
- to suggest ways of resolving these issues.

The new technologies can be used in various ways for the implementation of learning environments in ODL. One such learning environment [Koutoumanos et al., 1996] is abstracted in [Fig. 1] and [Fig. 2]. This learning environment will be used in the project. There are many subject matter areas in which one could experiment. Since all partners specialize in the area of Informatics, it was decided that EONT will be confined in this area. The partners are aware of the absence of any universally accepted standards for measuring effectiveness in education. This makes it very difficult to definitively substantiate claims about outcomes. Despite this fact, they decided to proceed, on the grounds that it deserves the effort.

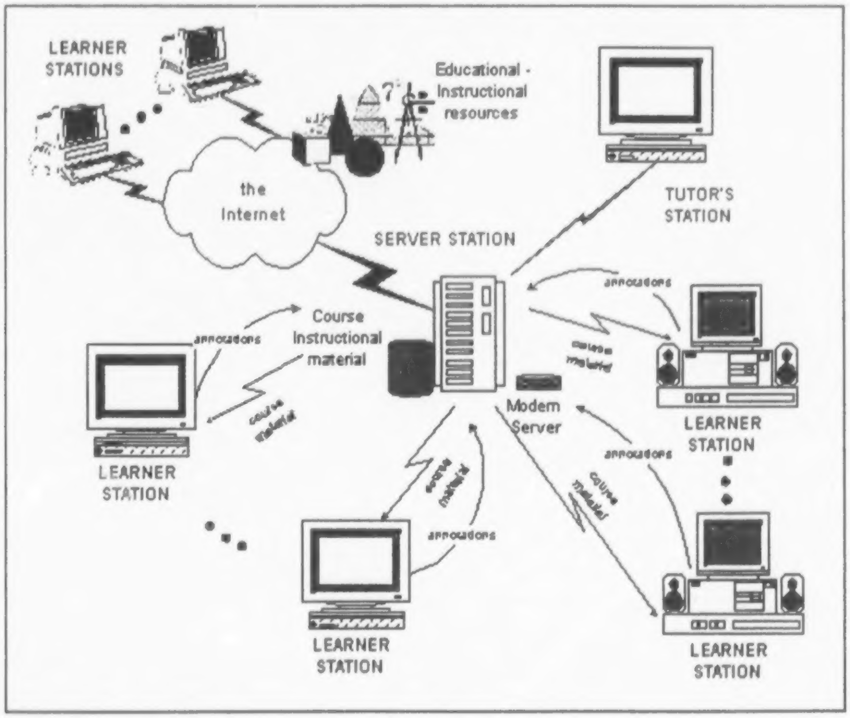

LEARNER STATIONS
Educational - Instructional resources
the Internet
TUTOR'S STATION
SERVER STATION
annotations
Course Instructional material
annotations
course material
course material
Modem Server
LEARNER STATION
LEARNER STATION
annotations
course material
course material
annotations
LEARNER STATION
LEARNER STATION

Figure 1: Schematic view of the learning environment to be used in the experiment.

BEST COPY AVAILABLE

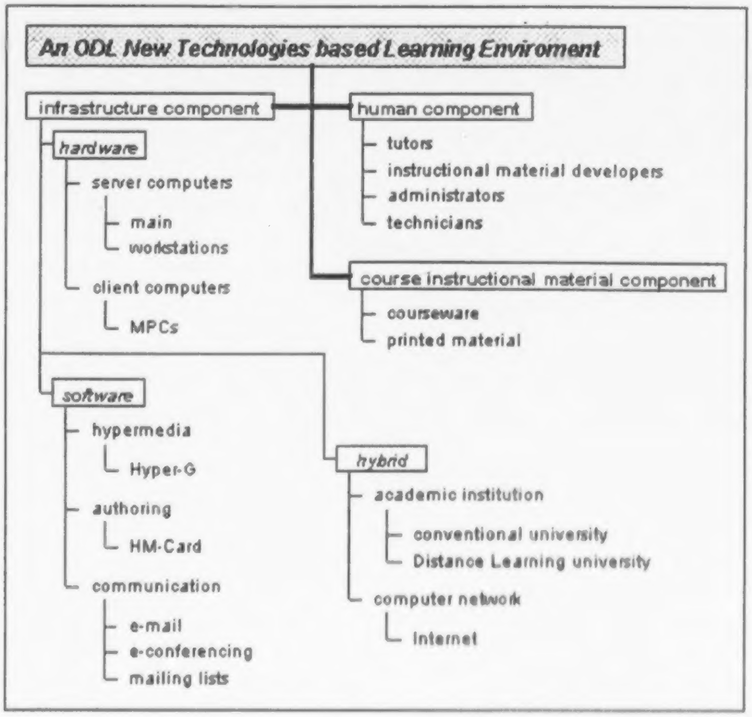

An ODL New Technologies based Learning Enviroment
infrastructure component
hardware
server computers
main
workstations
client computers
MPCs
software
hypermedia
Hyper-G
authoring
HM-Card
communication
e-mail
e-conferencing
mailing lists
human component
tutors
instructional material developers
administrators
technicians
course instructional material component
courseware
printed material
hybrid
academic institution
conventional university
Distance Learning university
computer network
Internet

Figure 2: Components of the learning environment to be used in EONT.

The abstraction depicted in [Fig. 1] constrains the learning environment of EONT to a networked environment, in which the instructional material is stored in a server computer. The learners access it through client multimedia computers, connected to the server via a computer network. The abstraction depicted in [Fig. 2] specifies the components of this learning environment. The hypermedia system HYPER-G [Flohr 1995, Maurer 1996] and the authoring tool HM-Card [Maurer et al. 1995] will be provided to the consortium by the Austrian partner who has developed them. Both possess innovative features, among which one should mention:

- powerful structuring mechanisms;
- access control and user identification;
- private and public annotations;
- sophisticated search mechanisms;
- multilinguality;
- interoperability with major systems on the Internet, including WWW [Berners-Lee et al. 1994, Vetter et al. 1994], Gopher, WAIS, telnet and ftp;
- client-server architecture, distributed over a computer network;
- accessibility from a variety of platforms (UNIX, X-Windows, DOS, MS Windows, Macintosh, etc.);
- hyperlinks not stored within documents but in a separate *link database*, allowing users to attach links to read-only documents that they do not own;
- the courseware can be easily extended and updated;

For the purpose of the experiment, it was decided that the learning environment will be used:

- as an alternative instruction delivery environment, for the distance learning universities; and
- as a supplementary instruction delivery environment, for the conventional universities.

Each partner will offer one course. These courses are shown in [Tab. 1]. The language of each course

BEST COPY AVAILABLE

will be the native language of the associated partner. Thus, the courseware will initially be in the native language of the partners, i.e. the national version. Subsequently, it will be translated into English, in order to have an international version for each course.

Partner	Course content	Language
P1	Introduction to Software Engineering	Greek (GR)
P2	Hypermedia Systems	German (DE)
P3	User Interface Design and Development	English (EN)
P4	Multimedia	Dutch (NL)
P5	Introduction to the Unix Operating System	Norwegian (NO)
P6	Elementary Course in Mathematics	French (FR)
P7	Software Engineering for Distributed Systems	German (DE)

Table 1: Courses to be offered.

4. Implementation Approach

The project started on December 1st, 1995, and will be completed in three years. The experiment will be performed in all seven universities. The learning environment will be the same for all partners except for the course instructional material. This will be developed by each university, as explained previously, in two versions -a national and an international one- and will consist of courseware and printed material. The national version will be offered once during the second year of the project and once during the third year. The international version will be offered only once, during the third year, and will be used for experimenting with the use of ODL as a means of providing learners in one partner's country with access opportunities to the course instructional material of the other partners. Data will be collected through questionnaires completed by the learners, from the respective live tutor, as well from the learning system itself. The obtained data will be analyzed separately for each partner and then collectively for all partners. From this analysis, a results report will be produced.

The course instructional material will be developed during the first year of the project and will be used for experimentation during the last two years. Care will be taken for developing courseware of high quality, since bad quality courseware would invalidate the experiment. An engineering approach will be employed for this purpose. A common methodology has been adopted by the partners, a sketch of which is abstracted in [Fig. 3]. It is based on a courseware development model and has as components: development methods, management methods, standards and tools. The *courseware development model*, depicted in [Fig. 4], will be a variant of the waterfall model [Marshall et al. 1995]. The figure indicates the main development subprocesses and the order in which these will be carried out. *Development methods* give answers to questions such as how the development will be carried out, as well as how to observe the proper incorporation of didactic and pedagogical principles in the courseware. *Management methods* provide solutions to managerial problems that will arise. *Tools* are used for supporting the development process. Finally, *standards* provide solutions to recurring problems.

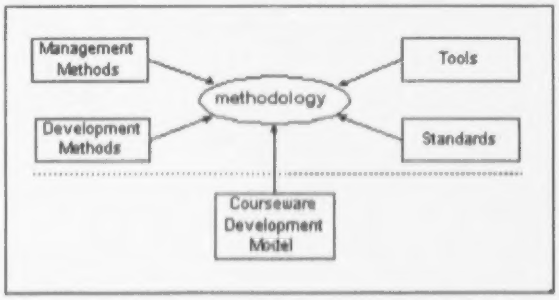


Figure 3: Methodology components for courseware development.

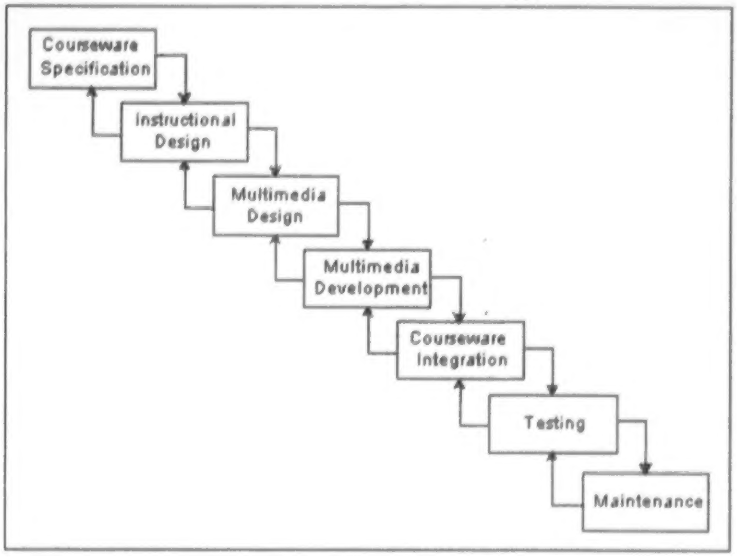


Figure 4: The courseware development model.

5. Concluding Remarks

The world of tomorrow will be digital. Information will be produced, stored and transmitted in digital form. Many tasks that are still performed manually will be performed automatically or semi-automatically. Automation will be the characteristic of tomorrow's information society. The new information technologies, especially computer networks and networked hypermedia systems, will play an important role in this evolution. It is certain that education will not be unaffected by this evolution. Many tasks that teachers perform today will be automated tomorrow and, as the automation of information processes has been very successful during the last decades, one can only expect that the same will happen with the automation of teaching processes.

ODL is expected to benefit the most from the evolution of new technologies. Although this sounds reasonable, it has to be proven by experimentation. This is the main objective of the EONT project, which addresses one particular ODL learning environment and one subject matter area. The findings of EONT have to be complemented with findings from other similar projects which will address different subject matter areas and possibly different ODL learning environments. Only through such a procedure can a reliable answer be found, concerning the problem of the effectiveness of ODL supported with new information technologies. Such a procedure is lengthy and expensive but the quality of education deserves all the effort and the associated cost.

References

[Berners-Lee et al. 1994]
Berners-Lee, T., Cailliau, R., Luotonen, A., Frystyk Nielsen, H., and Secret, A. (Aug. 1994). The World-Wide Web. *Communications of the ACM*. 76-82.

[Commission 1995a]
European Commission (September 1995). *SOCRATES Vademecum*. DGXXII Education, Training and Youth, Office for official publications of the European Commission. Luxemburg.

[Commission 1995b]
European Commission (September 1995). *SOCRATES: Guidelines for Applicants - Open and*

Distance Learning. Office for official publications of the European Commission. Luxemburg.

[Flohr 1995]
Flohr, U. (November 1995). Hyper-G Organizes the Web. *BYTE*. 59-64.

[Koutoumanos et al. 1996]
Koutoumanos, A., Papaspyrou, N., Retalis, S., Maurer, H., and Skordalakis E. (1996). Towards a Novel Networked Learning Environment. In *Proceedings of the World Conference of The Web Society (WebNet'96)*.

[Marshall et al. 1995]
Marshall, I. M. et al. (1995). The Mythical Courseware Development to Delivery Time Ratio, *Computers in Education*, 25(3). 113-122.

[Maurer et al. 1995]
Maurer, H., Scherbakov, N., and Scheider, A. (1995). HM-Card: A New Hypermedia Authoring System. *Multimedia Tools and Applications*, 1(4). 305-336.

[Maurer 1996]
Maurer, H. (1996) *Hyperwave: The Next Generation Web Solution*. Addison Wesley.

[Vetter et al. 1994]
Vetter, R., Spell, C., and Ward, C. (October 1994). Mosaic and the World-Wide Web, *IEEE Computer*. 49-57.

Acknowledgements

The project EONT is partially funded by the European Commission, Directorate General XXII, SOCRATES Programme, Chapter III, Action 2. The project's reference number is TM-OP-1995-1-GR-88 (1/0). The initial contract period is from December 1st, 1995 to November 30th, 1996. The contract is renewable for two more years, subject to an assessment of progress made and the demonstrated need for further support.

Appendix A: Structure of the Consortium

Partner	Institution	Country	Administrative person	Technical person
P1	Institute of Communication and Computer Systems (ICCS)	Greece	E. Skordalakis	A. Koutoumanos
P2	Institute for Information and Computer Supported Media (IICM)	Austria	H. Maurer	N. Scherbakov
P3	Katholieke Universiteit Leuven (KUL)	Belgium	H. Claes	E. Duval
P4	Open University (OU)	Gr. Britain	N. Wagstaff	D. Benyon
P5	Stord/Haugesund College (SHC)	Norway	H. Haugen	A. Staupe
P6	Université Lille I (UL)	France	A. Derycke	C. Vieville
P7	Fern Universität Hagen (FUH)	Germany	R. Bartz	B. Krämer

A Distance Education How-To Manual: Recommendations from the Field

Dr. Angie Parker, Ph.D.
Professor
Educational Technology
Gonzaga University
U.S.A.
aparker@soe.gonzaga.edu

Abstract: Distance education has opened the doors of education to literally thousands of students who could not otherwise receive instruction. Although the electronic highway is readily available to deliver the information, the issues of transition from traditional to distant instruction remain to be solved. The initiation of distance education must include the transition from teacher-centered to student-centered instruction, support of faculty, and the knowledge of how to restructure the coursework to meet the needs of the new world-cohort of learners.

Introduction

Distance education has opened the doors of education to literally thousands of students who could not otherwise receive instruction. While the advent of distance delivery has been important for students, the institutions producing the instruction have grappled with numerous issues for the purpose of providing high-quality education. This paper will look at three of these issues from the viewpoint of a distance educator.

The Retraining of Students

Since the advent of the Common School, students have come to accept the Socratic model wherein passive receipt of instruction, limited discussion and regurgitation of facts on examinations are commonplace. Distance education is substituting this timeworn paradigm with one that places far more responsibility on the learner. Students are now being asked to examine thinking and learning processes; collect, record, and analyze data; formulate and test hypothesis; reflect on previous understandings; and construct their own meaning (Crotty, 1994). The once teacher-centered instruction has now become student-centered and as a result, the once passive learner must now be engaged actively in creating personal knowledge that can be transitioned to new and different situations.

Although the paradigm has changed, students still arrive at the educational marketplace with the old expectations in place. Changing old ideas becomes the responsibility of the distant educator. This new cohort of students will require careful retraining if they are to comprehend the student-centered provisions of distance education. Not only will their focus become personally determined, but the depth of their learning and the degree of interaction will be seriously modified. The former teacher-centered model required little if any interaction from students. While the need to interact with others is an innate tendency that most students possess when they begin their education, the introduction of technology, which mediates interaction, often hinders that tendency. Therefore, it is essential that the retraining of students include brief technological instruction, coupled with periodic question and answer sessions to assure that the interaction will persevere and will include both pedagogical and social attributes.

Today's modern technology allows for interactivity between student and instructor to transpire in many formats. Teleconferencing, interactive video, conference calls, and CUSeeMe are only a few of the possibilities. It is important, therefore, for the interactivity to move beyond brief encounters between student and teacher and to invoke a feeling of cohesion between student and teacher, among the students themselves, and finally between the student and the material being studied. In the traditional setting of education, students interact easily and often to discuss course material as well as social topics. The same "chat" is important in the distance setting and must be carefully and richly sculptured by the instructor. Students who are unaccustomed to using a microphone to speak or the lack of visual contact with the person to whom the question is being addressed need to be instructed on how to utilize the technology to interact.

In addition to the social and pedagogical "chat" that is so important in successful distance education, is the student's ability to think critically about the topic being taught. McGiven (1994) believes success in distance education is directly related to the student's ability to critically analyze, to interpret information and to interact with peers to elaborate on the concept. Many students are unable to interact successfully until they have a basic understanding of how to analyze a question and how to synthesize the information being discussed. The advancement from lower level thinking skills to higher order skills is a most important issue that must be addressed during the retraining of students.

The changing needs of students suggest that distance education has the opportunity to foster a student-centered learning process and the responsibility to retrain students within the new model of instruction. This opportunity can be realized only when the new cohort of distance students is comfortable with the technology, with mediated interaction, and with the new paradigm of student-centered learning.

The Retraining of Faculty

Although advances in technology have allowed unique opportunities for the delivery of student-centered instruction to geographically diverse populations, the aggregate effectiveness of distance education ultimately rests with the faculty's endorsement of the new student-centered model of instruction. Most professors come to distance education with traditional teaching experience and find that the theoretical-based assumptions that worked successfully in face-to-face instruction do not translate well into technologically mediated instruction (Schieman, Taire & McLaren, 1992). As a result, faculty retraining is essential not only to assist with the use of the technology, but to also help with the revision of the instructional design.

The retraining should initially focus on moving the teacher from the podium to the sideline, from teacher-centered instruction to student-centered instruction. Once this objective is achieved, subsequent retraining components can encompass items such as the redesign of the syllabus, use of the technologies being employed and the integration of interaction in the course content. Reaching a comfort level for most faculty requires hours of rehearsal, trial and error, and planning. This is not a result of the difficulty with the technology itself, but instead a result of the mediated delivery. Lack of personal response and body language makes teaching at a distance a new experience. Additionally, the role of facilitator is unfamiliar to most faculty. Role playing and conceptualization of questions that require higher order thinking skills necessitate considerable time on the part of the distant educator.

Although often overlooked, the scheduling of the retraining is important as well. Orientation and training sessions should be scheduled well in advance of the beginning of the semester. Only then can faculty be expected to integrate the theories of the new paradigm with quality course content and, therefore, meet the challenges of distance education.

Redesign of the Course Content

The sentiment of many faculty entering the distance education arena for the first time is to teach the same course offered on campus with the addition of a few more handouts. To those who are experienced in the art of distance delivery, it is evident that the addition of a few more handouts is not the solution for successful course design. Recent work by McGiven (1994) found that the most important component in successful distant instruction was that of required and consistent interaction. Although much has been written about the need for interaction (Garrison, 1990; McGiven, 1994; Wagner, 1993) few researchers have offered specific ideas for integrating dialogue into distance education. Experience of the writer has indicated that the course design should focus on real world problems, students working in teams to find solutions, and consistent dialogue between class members and the instructor. Additionally, the questions posed should involve higher-order thinking skills such as evaluation, analysis, and synthesis rather than rote memorization. Once the question is' presented, either by the instructor or the students, hypothesizes should be discussed and finally teams of students should be assigned to explore possible solutions. In the early weeks of the course, forced interaction is usually required. This can take place by linking a portion of the final grade to the length and occurrence of interaction. In traditional instruction, this portion of the grade is usually referred to as attendance.)

After several weeks of successful interaction, the role of the instructor can shift to that of moderator. During this second phase, the instructor's input should constitute of no more than 20 percent of the total dialogue and should primarily involve setting the pace of the discussion as well as acting as the "provocateur". A well designed distance education course includes many topics for discussion, feedback from students as well as experts, and finally links to other sources of pertinent information. URL locations, literature citations and journal articles are only a few of the possibilities. Although most new distance instructors are primarily focused on dialogue and interaction within the confines of the course content, it is important to incorporate a degree of humor into the course. Humor can lighten the burden of the learning curve and generate a feeling of sincerity among those in the class.

Redesigning a course for distance delivery requires not only a rethinking of the content but also a new view of the learners. Opportunities for interaction, students working in distant groups, and dialogue constitute the major changes in course design. The syllabus must become the road map and the student become the driver when instruction is mediated by technology.

Summary

Distance education is made up of a network of learners and teachers who travel electronic highways and meet in virtual universities. This offers both a challenge and an opportunity. The opportunity lies in the access to education for a world-wide coalition of students. The challenge lies in the successful retraining of students, the support of faculty, and finally in the selection of subject matter and technologies that lend themselves to high levels of interaction and dialogue. Both the challenge and the opportunity serve to make distance education unique.

References

Crotty, T. (1994). Integrating distance learning activities to enhance teacher education toward the constructivist paradigm of teaching and learning. In Distance Learning Research Conference Proceedings, 31-37. College Station, TX: Department of Education and Human Resources Development, Texas A & M University.

Garrison. R. (1990). Quality and access in distance education: Theoretical considerations. In Theoretical Principles of Distance Education, ed. D. Keegan, 9-21. New York: Routledge.

McGiven, J. (1994). Designing the learning environment to meet the needs of distant students. Journal of Technology and Learning. 27(2), 52-57.

Dillon, C. & Walsh, S. (1992). Faculty: The neglected resource in distance education. The American Journal of Distance Education 3(2), 35-43.

Schieman, E. ,Taire, S. & McLaren, J. (1992). Towards a course development model for graduate level distance education. Journal of Distance Education, 7(2),51-65.

Wagner, E. (1993). Variables Affecting Distance Educational Program Success. Englewood Cliffs, NJ: Educational Technology Publications.

Relational Databases and the World Wide Web: Automatic Generation of Hypertext based on Reverse-engineered Meta Information

Richard Pönighaus, Johann Mitlöhner
Vienna University of Economics/Department of Applied Computer Science
A-1090 Augasse 2--6, Vienna, Austria

Email: poenigh@wu-wien.ac.at, mitloehn@wu-wien.ac.at

Abstract: In this work we describe an approach to the generation of HTML code based on meta information contained in typical relational database systems. In principle reverse engineering methods can automatically produce web pages from any database without the need for additional user input. Additional information can be supplied in order to improve the structure of the resulting web pages. We successfully applied our approach to several databases from different domains; in this paper, we use a database on classical records and its corresponding meta model to illustrate the method and point out problems which arise in automatic code generation from databases.

1 Motivation

With the ever-growing popularity of the World Wide Web people become used to exploring data in new ways. Navigating through hypertexts is becoming more and more common for users in diverse domains of application. The presentation of data in HTML format allows for plattform-independence and easy access for non-technical users.

Recently there has been growing reserach interest in meta-data [Metadata 1996] in diverse areas. Relational databases are capable of storing data in a highly structured way; when attempting to generate hypertexts from database records, the system catalog of the database can provide information on relationships between database tables on the meta-level. On the instance-level this allows for automatic creation of hypertext links between documents.

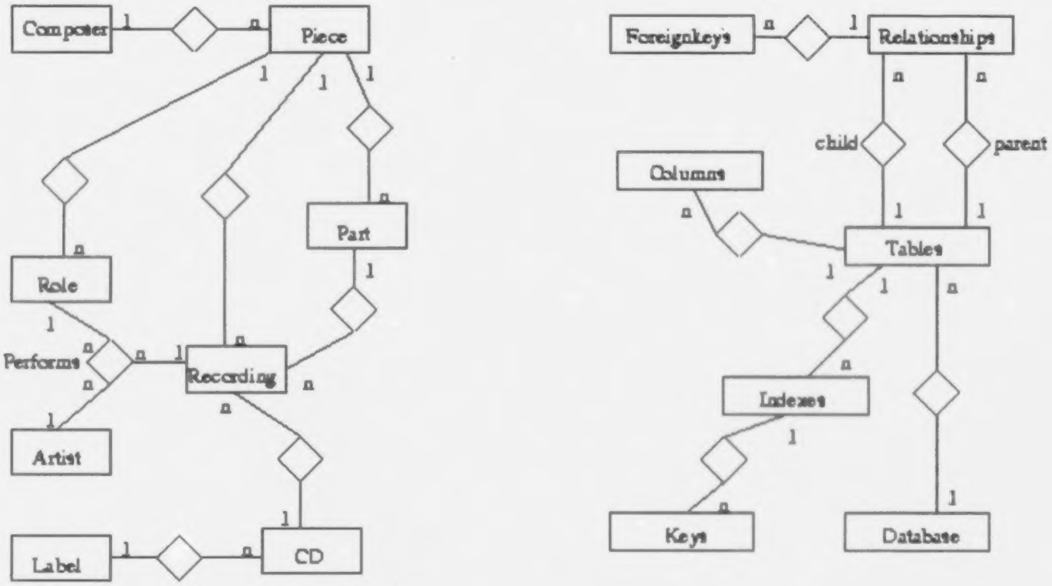
Among other things, database systems offer the advantage of being able to store large amounts of data in a consistent way. Today, many database applications exist in a variety of application domains. Several reasons exist for making data stored in these applications accessible via the web:

- Non-technical users typically only have restricted or no access to RDBMS (Relational Database Management System) data, since the interface provided often requires database knowledge the average user does not have. Therefore, making available the huge amounts of data which exist in database applications in a way all users can instantly grasp offers large benefits to organizations. By simplifying the access to the data users who lack skill or time to use databases directly can still apply information systems in their work.
- Generating hypertext from different RDBMS allows for a unified view of information: material from diverse database systems can be combined into a single hypertext, eliminating the need for users to switch between systems when accessing data.
- Hypertext generated from database systems which do not support multiple users can be placed on web servers and then be viewed by many users at once, resulting in effectively turning the application into multi-user mode (at least for browsing).

Those points are valid not only for Intranets (within an organization), but also for Internet users distributed all over the world. In the next sections we demonstrate the process of automatic hypertext generation by using the example of a database of classical records.

2 The Music Database

In our model a piece of music is composed by exactly one composer. Several artists may take part in the recording of a piece, optionally performing in a role specific to the piece (such as singers in opera). The record may be of a part of the piece, or of the whole piece. A CD may contain one or more recordings.



(a) (b)

Fig. 1: The ER-diagramms for the records database (a) and the system catalog (b)

The relevant entities are Composer, Piece, Part, Artist, Role, Recording, CD, and Label; the cardinality of relationships between those entities are one-to-many, except for the many-to-many relationship Performs. The corresponding ER-diagram is shown in [Fig. 1a] (the ER-modelling approach was first introduced by [Chen 1976]).

How can this database be viewed? For a hypertext, we need a starting point. In the example, a good starting point would be a list of composers. Following a link for a specific composer would answer the question `` *What are the pieces written by that composer?*". Following the link for a piece would show a list of recodings of that piece. The link of a recording would lead to the CD that contains that recording.

Another starting point would be a list of the artists. Following a link on an entry in that list would answer the question `` *What recordings did this artist take part in?"*.

In the following sections we will discuss our approach to the automatic generation of HTML code from a relational database. We will draw on examples from the music database described in this section.

3 Significance of Reverse Engineering

When building applications, we must often start from old system, and we often do not exactly know their structure, as has been summarized by [Waters & Chikofsky 1994]:

> ``... while many of us may dream that the central business of software engineering is creating clearly understood new systems, the central busines is really upgrading poorly understood old systems. By implication, reverse engineering is arguably one of the most important parts of software engineering, rather than being a peripheral concern."

Relational system are in principle very well suited for reverse engineering. [Markowitz & Makowsky 1990] show a theoretically sound approach but assume good design and full normalization which often does not occur in real applications.

```
CREATE TABLE Composer                          CREATE TABLE Piece
       (ComposerID   INTEGER NOT NULL,               (PieceID      INTEGER NOT NULL,
       Name          CHAR(30) NOT NULL,                ComposerID   CHAR(18),
       FirstName     CHAR(30),                         Title        CHAR(50) NOT NUL
       Born          DATE,                             PRIMARY KEY (PieceID),
       PRIMARY KEY (ComposerID));                      FOREIGN KEY by (ComposerID)
                                                          REFERENCES Composer
                                                          ON DELETE SET NULL);

CREATE UNIQUE INDEX Composer_ix                CREATE UNIQUE INDEX Piece_ix
  ON Composer (ComposerID ASC );                       ON Piece (PieceID ASC);
```

Data and index definition for the tables Composer and Piece in SQL

Taking a look at the SQL statements in the music database, we instantly see possibibilities of automatic link generation. The table Composer is referenced by the table Piece; obviously, this reference can be used in generating hypertext link. Assuming we translate each table entry into a web page, one can easily imagine how for each piece of music in the database a link can be added to the field ComposerID in table Piece. The SQL statement defines this field as a foreign key which references the table Composer.

Unfortunately, SQL statements for table creation are not normally available in legacy systems. However, the information is not lost: the RDBMS automatically manages the corresponding structure. In the following section we take a look at the way a RDBMS is storing meta-data about applications.

4 The System Catalog

RDBMS store meta information about applications in system catalog tables. The structure of all objects derived from the data definition language (DDL: SQL statements like CREATE TABLE, CREATE INDEX) in the user databases are described in the catalog tables. Sophisticated systems store other information as well, e. g. on permissions, statements embedded in application programs (DML). In [Pönighaus 1995] an analysis on the usage of SQL statements based on catalog data was performed; here we concentrate on DDL. The meta information contained in the catalog tables is accessible via standard SQL statements. [Fig. 1b] shows an ER-model of a simplified system catalog (our meta base).

System catalogs vary somewhat between database systems. In the following we describe relevant parts of our system catalog here; most widely used database systems provide the information needed, although the corresponding tables and fields would be named differently.

Database:

this table contains the names of the databases and their creators. In a typical RDBMS, a user can own more than one database.

Tables:

this table contains the names of the tables which form the application database. Entries in this table consist of table name, database name and creator name.

Columns:
this table describes the columns of all application tables. An additional Remarks field can be used for the web page generation if any text has been supplied by the table designers.

Relationships:
this table contains information on tables that are referencing other tables (or itself). It states the referencing table (Tbname), the table being referenced (Reftbname), and the name of the relationship.

Foreignkeys:
relationships to other tables are described here, for each relationship contained in the catalog table Relationships, the columns involved in the relationship are stated here.

Indexes, Keys:
these tables define the indexes created on application tables and the columns involved.

Let us take a look at how the SQL statements from the music database application are translated by the RDBMS into system catalog entries. If the database Music has already been created by the user Smith the catalog table Database contains (among others) the following entry:

```
DBname | Creator
----------------
 ...   |
Music  | Smith
 ...
```

The creation of a new table results in an entry being added to the catalog table Tables:

```
Tbname   | Type | DBname | Creator
----------------------------------
 ...     |      |        |
Composer | E    |  Music | Smith
 ...
```

For each columns of the newly created table, an entry is added to the catalog table Columns (for simplicity we omit the creator name in the following):

```
Colname    | Tbname   | Colno | Coltype
---------------------------------------
 ...       |          |       |
ComposerID | Composer |     1 | INTEGER
Name       | Composer |     2 | CHAR
FirstName  | Composer |     3 | CHAR
Born       | Composer |     4 | DATE
 ...
```

The primary key declaration and the index creation cause the tables Indexes and Keys to contain the following entries:

```
Ixname      | Tbname   | Uniquerule
-----------------------------------
 ...        |          |
Composer_ix | Composer | P
 ...
```

```
Ixname      | Colname    | Colno | Colseq | Ordering
---------------------------------------------------
 ...        |            |       |        |
Composer_ix | ComposerID |     1 |      1 | A
```

In the application table Piece the field ComposerID is a foreign key referencing entries in the table

Composer. The declaration ``FOREIGN KEY..." means that the entries in the column ComposerID relate to entries in the Table Composer; the relationship has been given the name ``by". This relationship name is entered into the catalog table Relationships by the RDBMS:

```
Tbname | Relname | Reftbname | Colcount | Deleterule
-----------------------------------------------------
 ...   |         |           |          |
Piece  | by      | Composer  |        1 | C
 ...
```

To keep track of the columns taking part in the reference, entries into the catalog table Foreignkeys are added. In this case, there is only one corresponding row:

```
Tbname | Relname | Colname    | Colno | Colseq
----------------------------------------------
 ...   |         |            |       |
Piece  | by      | ComposerID |     2 |      1
 ...
```

In the following section we describe how we make use of the catalog data discussed above in the generation of HTML code.

5 Overview of our approach

We use the relationships between database tables which can be found in the system catalog of the RDBMS to automatically build a hypertext from a given database. The resulting HTML code can be viewed with a standard web browser.

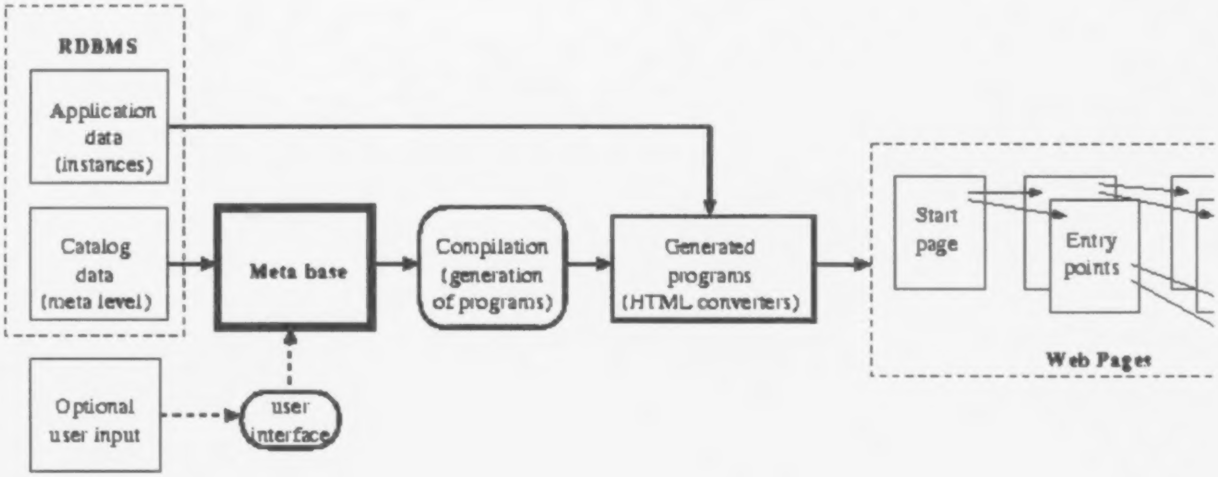

RDBMS
Application data (instances)
Catalog data (meta level)
Meta base
Compilation (generation of programs)
Generated programs (HTML converters)
Start page
Entry points
Web Pages
Optional user input
user interface

Fig. 2: An overview of the web page generation process

[Fig. 2] shows an overview of our approach. We transfer the data from the system catalog to a meta base; ideally, this data is sufficient for the hypertext generation. However, the user can supply optional additional input in a format consistent with the catalog data. Additional user input allows for increasing the readability of the resulting web pages and for incorporating additional features into the hypertext as explained later.

BEST COPY AVAILABLE

Our program generator is written in the standard Unix record processing language AWK; the programs generated by our system are in AWK as well. We chose this language for our prototype system since it is very well suited for this type of processing, and it is widely available in commonly used operating system and hardware environments. The meta base is currently implemented in MS ACCESS; additional user input is possible via forms. From the meta base we generate a set of programs. After exporting the application data from the database system those programs create corresponding web pages. Links to other documents, meaningful section titles and formatting commands are added by those programs.

The generated programs are run to produce the web pages from the current application data. A start page is automatically generated to provide a starting point for the data exploration. The resulting web pages form a three-layer hierarchy: the start page, the pages for the different entry points, and the pages for the data instances.

6 Hypertext Creation

Our hypertext documents are composed of the entries in the database and links to corresponding other entries. In the following, we discuss basic types of documents and special cases.

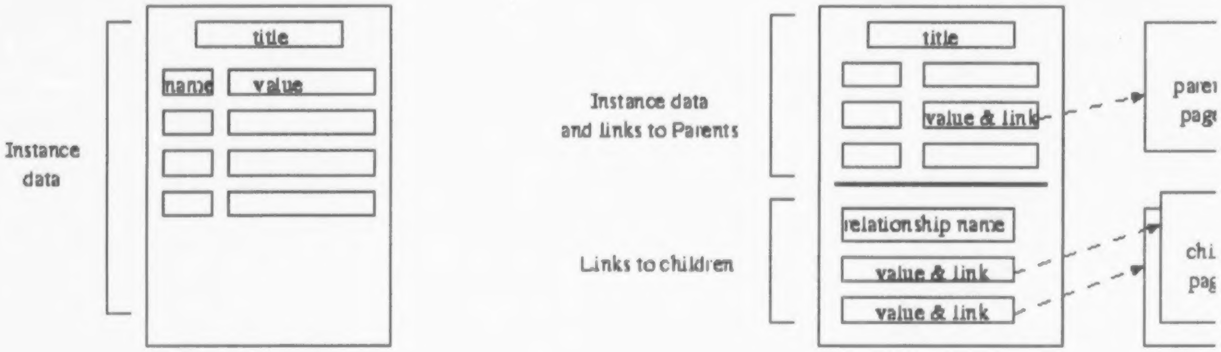
6.1 Document Types

``Stand-alone" document:

> In the simplest case of a table containing no references to other tables and not being referenced by any other table, each entry in the database (a record in a table) is translated into a HTML file containing a title, the names of the fields and their values. A sample document is shown in [Fig. 3a]. If any label text has been supplied in the catalog table Columns or by additional user input, this text is used as the label for a field; otherwise, the column name is used for this purpose.

References between tables:

> This more interesting case is depicted in [Fig. 3b] which shows a web page consisting of two parts separated by a rule. Consider a web page for a piece of music: the upper part contains instance data and links to parents. A parent table is a table that is referenced by another table:
>
> the second field in the upper part (ComposerID of table Piece) of Fig. 3b] contains a link that points to a parent (an entry in the table Composer), in this case the composer of the piece. A link to the corresponding web page is provided together with the value (possibly expanded, see sec. 6.2).
>
> The lower part shows a list of children of this piece, i. e. a list of recordings. We use the expanded keys to represent the children. Note that in the case of 1:n-relationsships a parent can have several children.



(a) (b)

Fig. 3: A generated stand-alone document (a) and a document with references (b)

When automatically generating hypertext from database data, we encountered several problems which seem to persist across different application domains. We show those of the problems we have solved in the following sections.

6.2 Key Expansion

Consider the table Piece: each entry references an entry in the table Composer via the field ComposerID. In many cases this would be a unique number for each composer (an artifical key). In the table Piece this key is contained in the referencing field. Obviously, the resulting web page would not look satisfying, since seeing e. g. the number 42 in the field ComposerID does not tell much; the user would have to follow this link to see what the number stands for.

Substituting the composer's name for the ComposerID would be a substantial improvement here. We saw no way of automatically deducing which fields of a referenced table are meaningful for the reader of the web page. This information has to be supplied by additional user input in the form of additional indexes on tables; existing alternate unique indexes can help the user in this process, since these indexes usually are created on fields that describe an entry in a way meaningful to a human reader. By entering additional indexes for tables the user can provide the HTML generator with information on which fields are ``natural keys" for that table i. e. the contents of those fields will identify the entry to the user; e. g. in the table Composer, natural keys would be the columns which contain the composer's first and last name. To give this information, the user would enter an additional index of type ``A" (alternate). The generator uses this information for the key expansion. This information is also used to create more meaningful titles of web pages in the form of <entity name> : <alternate key>.

In complex cases, the key expansion is done in several steps: the meta base table Relationships describes a network of connections between tables in the application which is traversed by the generator using a graph algorithm until it arrives at meaningful contents for hypertext fields.

6.3 Associative Entities

In the music database, the information on which artists participated in the recording of a piece is stored in a table Performs. This table is an associative entity: it is neccessary for the implementation of the many-to-many relationship: more than one artist can take part in a recording, and each artist can take part in more than one recording in a particular role. However, the table Performs is not interesting in itself; the reader of the hypertext would prefer to perceive the information only in the form of a list added to the Recording page: this list should directly state the artists taking part in the recording plus their role, not the entries in the table Performs that reference this recording. The table Performs should not be part of the hypertext system in the same way that Composer, Piece, Artist, Role and Recording are: we want the information of artist participance in recordings, but we do not want a web page for each entry in the table Performs.

If foreign keys are defined in the application, we can automatically deduce the information of whether an entity is an associative entity and should be omitted: in our prototype no web pages are created for tables that consist only of foreign keys.

Often information on foreign keys is not supplied within the RDBMS; our experience confirms what [Premerlani &Blaha 1994] say: ``Finding candidate keys is easy, finding foreign keys is difficult."

The declaration of foreign keys is often lacking for reasons of performance, or because legacy RDBMS did not support the corresponding definitions. Additional user input (deriving from knowledge of the application domain) can clarify the situation.

6.4 Groupings and Listings

Readability and accessability of our generated hypertexts have been significantly improved by the introduction of groupings and listings which serve as entry points to sets of instance web pages [Fig. 2]. Hierarchical groupings and ordered lists are defined by special index types (indicated in our meta base by ``G" and ``L", resp.). These have to input by the user.

7 Future Work

During the development of our system, several topics have been postponed to future investigation. They fall into the two areas of formatting and structural improvements:

7.1 Formatting Improvements

Treatment of long lists:

in cases where an entry in a parent table is referenced by many children, the list of the relationship can grow very long: in the music database, the composer page for Wolfgang Amadeus Mozart would list several hundred corresponding pieces. This list would be somewhat cumbersome to view in a single web page. An intermediate directory should be generated in cases where lists grow to long, containing links such as ``A-B", ``C-D"... Also, this could in many cases significantly reduce the time required for loading parts of lists, which is especially important in the case of low Internet transmission rate.

Very short pages:

in a database, some tables can contain only very few short columns. However, presently a web page will be generated for each table entry, unless it is an associative entity (see Section 6.3). This is not satisfying; a better solution in this case would be to generate web pages containing more than one table entry, separated by a distinguished border, such as a horizontal rule.

Code Translation:

In many applications short strings or single letter codes are used to code information (e. g. in the Music database, ``CH" stands for choir, ``S" for soloist). An automatic translation of such strings would improve the readability of the hypertext. Adding a table Codetranslations (one-to-many relationship to Columns) to the meta base would implement this concept.

7.2 Structural Improvements

Weak entities:

in our system catalog, the table Columns would be a weak entity: it cannot exist without its parent. All information in entries in the Columns table should be added to the strong entity Tables. An analysis based on 3,000 tables in a large Austrian bank showed an average of 13 columns per table [Pönighaus 1993]: this means that if we generated a system catalog browser we could save 13 web pages per entry in the table Tables (3,000 instead of 3,000 + 39,000 pages).

Denormalized data:

for performance reasons, data in practical applications are often denormalized: e. g. many-to-many relationships are implemented via fixed-length arrays [Premerlani &Blaha 1994]. In order to generate appropriate hypertext this situation must be recognized (by additional user input) and handled accordingly.

Slices:

[Isakowitz et al. 1995] introduced the concept of information slices in their Relationship-Management-Methodology (RMM): splitting an entity into meaningful slices for presentation of information, resulting in a separate web page for each slice of a table entry. In the present version of our generator, all fields of a table would be presented in one page. There are situations where this is not satisfying; we consider the integration of this concept into our meta base to be comparatively easy.

Dynamic HTML generation:

in our prototype the generation of the hypertext is done in batch mode, but the meta base can also be used for dynamically generating SQL statements which extract the relevant data from the RDBMS. In this case the data would always be up-to-date, which obviously is a requirement in many applications. The concepts discussed in this work essentially remain unchanged for this task; in particular, there will be no need for any changes to the meta base.

8 Conclusion

In this work we show an approach to the automatic generation of hypertext from relational databases using metadata. Reverse engineering applied to relational databases proves to be applicable to this problem, since the referencing information provided by the database system is ideally suited for automatically generating hypertext links between documents for instances of entities. The use of a meta base which is extracted from the system catalog and which can be augmented by the user allows for a flexible solution to the problems typically found in practical database applications. In this text the method is applied to a database on music records. Interested readers are invited to take a look at the music database at http://exaic.wu-wien.ac.at/~poenigh/platten/entry.htm.

References

Chen 1976

P. Chen. The Entity-Relationship-Model -- Towards a Unified View of Data. *ACM Transactions on Database Systems*, 1(1), March 1976.

Isakowitz et al. 1995

Tomas Isakowitz, Edward A. Stohr, and P. Balasubramanian. RMM: A Methodology for Structured Hyermedia Design. *Communications of the ACM*, 38(8):34--44, August 1995.

Markowitz & Makowsky 1990

V. M. Markowitz and J. A. Makowsky. Identifying Extended Entity-Relationship-Structures in Relational Schemas. *IEEE Trans. Softw. Eng.*, 16(8), August 1990.

Metadata 1996

Proceedings of the First IEEE Metadata Conference. April 16-18, 1996, Silverspring, Maryland (http://www.nml.org/resources/misc/metadata/proceedings/meta_home.html).

Pönighaus 1993
Richard Pönighaus. *Performance-Prognosen für relationale Datenbankmanagementsysteme*. PhD thesis, Vienna University of Economics and Business Administration, Vienna, Austria, 1993.

Pönighaus 1995
Richard Pönighaus. Favourite SQL Statements -- An Empirical Analysis of SQL Usage in Commercial Applications. In Subhash Bhalla, editor, *Informations Systems and Data Management/Lecture Notes in Computer Science, number 1006*, Berlin, 1995. Springer.

Premerlani &Blaha 1994
William J. Premerlani and Michael R. Blaha. An Approach for Reverse Engineering of Relational Databases. *Communications of the ACM*, 37(5), May 1994.

Waters & Chikofsky 1994
Richard C. Waters and Elliot Chikofsky. Reverse Engineering -- Progress Along Many Dimensions. *Communications of the ACM*, 37(5), May 1994.

Johann Mitlöhner, Richard Pönighaus, August 1996

The MARBLE Project: A Collaborative Framework for Educational Courseware Design

Sarah Price (sprice@icbl.hw.ac.uk), Heriot Watt University

Patrick McAndrew (patrick@icbl.hw.ac.uk), Heriot Watt University

Roger Rist (roger@icbl.hw.ac.uk), Heriot Watt University

Terry Mayes (jtma@glasgow-caledonian.ac.uk), Glasgow Caledonian University

Eric Bonharme (e.bonharme@dcs.napier.ac.uk), Napier University

Ray Land (r.land@central.napier.ac.uk), Napier University

Mary Cuttle (mcuttle@ed.ac.uk), University of Edinburgh

Jeff Haywood (J.Haywood@ed.ac.uk), University of Edinburgh

Hamish MacLeod (Hamish.Macleod@ed.ac.uk), University of Edinburgh

Abstract: The MARBLE Project is a collaborative venture which provides on-line resource-based learning materials to students at three Higher Education Institutions in Scotland. As *per capita* funding from central government decreases without a decline in graduate quality, new methods are sought in Universities to cope with this challenge. This paper discusses the collaborative beginnings and implementation of MARBLE, and then focuses specifically on one of the ten components of the range of on-line courseware. Since one influential report into Higher Education in the early 90s claimed that "Duplication of effort is wasted effort", we assess to what extent MARBLE's collaborative nature causes efficiency gains and yet, crucially, still enables the student to advance in the learning process.

This paper describes the MARBLE Project, which is an innovative, collaborative attempt to use the power of the Metropolitan Area Networks (MANs) in Scotland for effective courseware design and dissemination across the World Wide Web. The MANs have provided the technical motivation for MARBLE: each offers FDDI at 100 megabits per second and ATM at 155 megabits per second. The four MANs are currently linked via the high speed SuperJANET academic network, which provides internet links, and will shortly be fully interconnected.

The background for the creation of MARBLE is the changing pattern of Higher Education in the UK during the last decade. In brief, the major influences for change are increases in student numbers without *per capita* increases in funding, along with a greater external scrutiny of the quality of teaching and learning with the expectation that graduate quality will be maintained or even improved.

In an attempt to evaluate these changes and steer a way forward through this challenging situation in Higher Education, an influential report written by a committee of Scottish University Principals in 1992, concluded that:

"Computer Based Learning may not offer just another approach to learning, but may, in fact, be crucial to education's ability to meet the needs of the next few years"

[MacFarlane, 1992, p85]

Strategies were proposed and implemented shortly thereafter which led to the creation of the MANs and to a vision of the distribution of computer-based learning materials becoming reality. MARBLE is one project which contributes to this process.

The collaborative nature of the Project is also commensurate with the changed climate. Creating effective multimedia or hypertext courseware can be costly in terms of time and money; it is also extremely easy to do it badly. The notion of individual lecturers preparing their own educational software for their own immediate teaching needs has become increasingly viewed as less than ideal, given the necessity of constraining costs and avoiding duplication of effort wherever possible. MacFarlane foresaw that an "unprecedented degree of collaboration and sharing between different institutions" was required, and noted that "Duplication of effort is wasted effort" [MacFarlane, 1992, p33]. Diana Laurillard suggested further in 1993 that the problem could be addressed by producing educational software by consortia:

"Because of the importance of collaboration for greater efficiency of production of materials, and the greater likelihood of widespread dissemination of them, any central funds made available for development of these materials should be given only to collaborative consortia"

[Laurillard 1993, p230]

The Funding Councils in the UK have moved some way towards adopting such a model, such that the current Teaching and Learning Technology Programme (TLTP) required that all bids for funds to produce coursework software involved at least three institutions as members of a consortium.

The infrastructure for widespread dissemination of web-based courseware is firmly in place in Scotland (the high capacity MANs) and the time is right for such a collaborative venture (the technology-friendly direction taken at the highest level in response to the demands made of Higher Education). With these two essential building blocks in place, the MARBLE Project proposed a matching of suitable academic subject areas with identified university departments who were willing to tread new ground and incorporate some of the most innovative web-based capabilities into their curricula.

A Collaborative Framework

An essential feature of the MARBLE framework is that it is the collaborative product of three universities, and bridges subprojects in several cognate subject areas. A Project Officer is employed at each university to provide the technical and organisational input; a Project Management team of six consists of permanent staff from the three sites; and seconded academics from the university consortium provide the direction for the subprojects, plus the core academic knowledge upon which the system depends.

The 10 subprojects contained within MARBLE cross a range of subject areas. As even the most impressive new technology will not address the problems of the late 90s teaching environment unless the teaching staff are won over by it, MARBLE has attempted to "seed" the use of the MAN for resource-based learning across the following subjects:

Biological Sciences Databases

The Built Environment

Computing in Teaching and Learning

Geotechnical Engineering

Interactive Vision

Introductory Statistics Teaching in Psychology

Library Information Retrieval

Mathematical Assessment

Molecular Sequence Databases

The WWW for Computing Courses

Thus, many academic staff will find resource-based learning materials in fields not too distant from their own, which they can clone and adapt to their own teaching with minimum effort.

Convincing the Teachers

Most of those involved in the process of designing educational software are aware of the potential for distrust, or, at best, unease, on the part of some course tutors towards such work. One of the most effective ways to allay these fears and reiterate that computers can be viewed as a tool to reduce workloads and improve quality is to involve them with the technology and show that it can be beneficial. The MARBLE Project has attempted to demonstrate the new technology to the very people who must be won over by it if the changing ratios in Higher Education are to be coped with effectively. The use of the fastest multimedia communication techniques available, as described above, helps expand the skills of several staff in the universities, and to explore for themselves the potential and usefulness of these techniques in their everyday work.

The Core Problems

The core problems associated with effective distributed courseware design can be subdivided into three essential strands. One strand represents the problem of designing material that is both conceptually demanding enough to be useful in H.E, and yet general enough to allow specific tailoring by tutors to suit individual curricula, levels of student knowledge and methods of delivery. Another strand is concerned with achieving effective collaboration over a distance between disparate institutions and departments with no particular collaborative history. The third strand is concerned with how to make material provided on the Web more than merely expository, or more than a straightforward translation of material from a paper environment to html. In other words, the solution to the problem posed by the third strand requires one to use techniques of course design which engender and facilitate deep learning.

Insights on the first strand of achieving a flexible courseware framework have been provided by the INTERACT Project, a collaborative venture between Heriot Watt and Strathclyde Universities, Scotland, and the University of Cambridge, England, which sought to create a simulation environment in an engineering context. The key re-usability of resources concept was ably demonstrated ably through the generative nature of their *Interact Simulation Environment* which used the same generic simulation in a range of four different educational contexts, by different tutors and with students of different levels of ability. (See References: INTERACT)

MARBLE has approached the need to secure effective collaboration - the second of the three core problems - by employing the most powerful electronic communication tools at our disposal, facilitated by the ample capabilities of the Metropolitan Area Network. We are currently experimenting with such facilities as desk-top video conferencing and HyperNews, a dynamic messaging program that runs within any web browser. The MARBLE Project breaks new ground further in this regard as not just one academic area is broached by the collaborative team, but ten.

Maintaining standards on a decreased budget

The third problem identified in the list of three core problems above, concerns how to build into one's courseware the attribute of enabling students to achieve deep learning. This problem is actively confronted by the MARBLE subproject entitled Computers in Teaching and Learning. We shall focus on this sub-project in greater detail to assess how some of the objectives above have been realised.

Computers in Teaching and Learning (CTL) is a component of an MSc course (Human Computer Interaction) which makes full use of the World Wide Web as a learning resource and encompasses much more than just making lecture notes available on the Web (though notes with suitable hypertext links are indeed there). A Document Type Definition (DTD) specifies an SGML (Standard Generalized Markup Language), which recognises the different components of the course: tasks, articles, bibliographies, on-line exam questions, etc. This means that web materials can be created with the emphasis on content, rather than style, the latter being guaranteed always to be consistent and syntactically correct. It is an especially effective way to achieve clear hierarchical organisation of a large web site, but most importantly the output can be directed in more than one way, thus ensuring that resources are reusable and flexible. The same marked-up material can generate HTML for viewing on the Web, and it can also translate the source material into *Latex* and then on to paper-based materials, if required. The implementation of this MARBLE component has implications for the re-usability of resources issue. It would be possible to empty out the CTL specific contents from the shell structure and to refill it with material from other academic disciplines which could benefit from being presented in a similarly structured way, e.g., Politics, History, Sociology.

The syllabus is task-based and the tasks are clearly explained to students at an appropriate URL, along with the dates upon which the completed assignments are expected to be electronically delivered. This is an appropriate structure in that most educational psychologists would agree that an approach based simply on delivering content - without a framework of tasks in which knowledge is *constructed* as well as passively received - stands a high chance of failing.

Innovative Use of Resources

The shift to mass higher education has meant that valuable, meaningful dialogue between tutor and student has necessarily become a less frequent occurrence, and yet, as Laurillard points out, dialogue remains at the heart of the educational experience, [Laurillard, 1993]. Formerly, the student would be exposed to new concepts in an expository way through lectures or books, and after a reasonable (and possibly unprofitable) time, struggling with the new information, would have the opportunity of engaging in mutual questioning and reflection with a tutor sensitive to the context. With the assignment of relevant tasks to consolidate the material, and the correction of misleading perceptions by the tutor, eventually full understanding would dawn.

One way in which the CTL course breaks new ground is through the resources available to the students in executing these tasks, and specifically in how the computer enables them to achieve the essential dialogue and reflection which current thinking suggests facilitates learning at a deep level. Dialogue can take place in CTL in the time-honoured way with the tutor, with the computer (see below) or in a way which is gaining more and more interest as an effective teaching method: through peer tutoring. As one of the course tutors states:

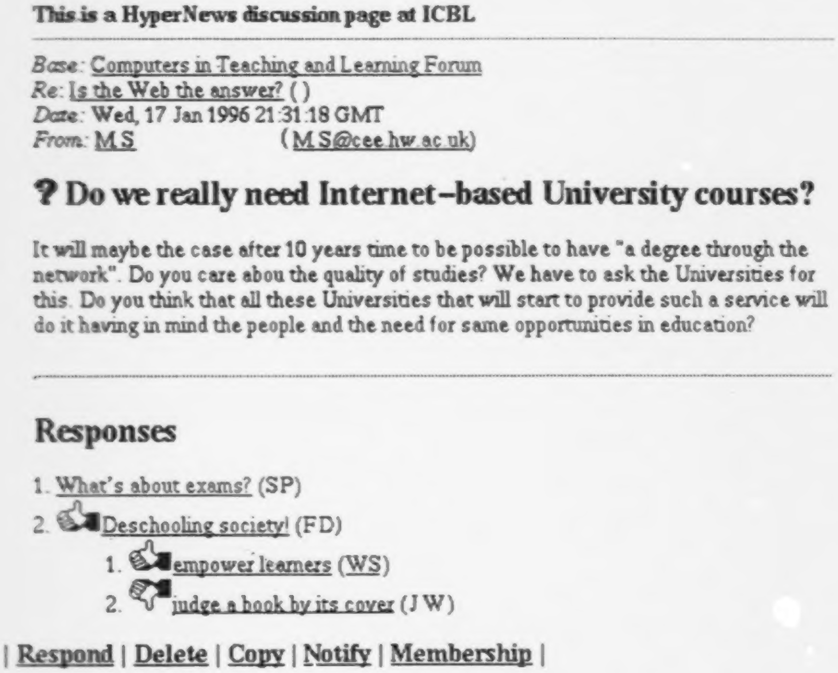
"Learning is often most effective where there is explicit peer tutoring; there is good evidence in favour of co-operative learning from studies comparing teacher-versus-peer mediated instruction. Some of the research has found that one to one peer tutoring is even more effective than supplementary instruction by a teacher in a small group setting. Not surprisingly, studies have also shown peer-tutoring to be highly cost-effective."

[Fowler et al, 1996]

The assigned tasks, mentioned above, are specially designed to encourage relevant group interaction. The means of achieving dialogue with peers can be through HyperNews, Global Chat, or through a system devised at ICBL to allow students to publicly annotate information disseminated as part of the course, as if they had a paper copy of the information and were leaving comments in the margins for others to see. Though electronic mail is also available, the value of the many-to-many group dialogue is that it is a forum for helpful discussion between persons having the same sort of learning experiences, facing potentially the same sorts of problems in comprehension. Peer tutoring is particularly helpful in broadening and deepening the learning experience.

Fig 2 shows an extract from a HyperNews session which runs through Netscape. It resembles a normal bulletin-board type News forum, but operates for class members only. Each assigned task may spark off several threads of related discussion within the group. Most of the input to the forum is initiated by the students themselves, but the occasional thread is authored by one of the course tutors who monitor the discussion in the event that expert help is needed to resolve a problem or simply to offer an alternative explanation. All group discussion which takes place as part of the course is captured for use by future students, or for propagation in the Answer Garden (see below) or for tutors as essential feedback on how course elements are received and understood.

Experiments have also been carried out on this component of the MARBLE Project on putting the *Answer Web* tool in the hands of the learners themselves. The concept of the *Answer Web* grew out of the Answer Garden, an idea first developed by Ackerman and Malone [Ackerman & Malone, 1990] and allows the development of databases of frequently asked questions (FAQs) that grow "organically" as new questions and answers arise. (The extension of the *Answer Garden* into a web-based tool for the INTERACT Project - the *Answer Web* - is discussed in [Smeaton & Neilson, 1995])



This is a HyperNews discussion page at ICBL

Base: Computers in Teaching and Learning Forum
Re: Is the Web the answer? ( )
Date: Wed, 17 Jan 1996 21:31:18 GMT
From: M.S (M.S@cee.hw.ac.uk)

? Do we really need Internet–based University courses?

It will maybe the case after 10 years time to be possible to have "a degree through the network". Do you care abou the quality of studies? We have to ask the Universities for this. Do you think that all these Universities that will start to provide such a service will do it having in mind the people and the need for same opportunities in education?

Responses

1. What's about exams? (SP)
2. 👍 Deschooling society! (FD)
    1. 👍 empower learners (WS)
    2. 👎 judge a book by its cover (J W)

| **Respond** | **Delete** | **Copy** | **Notify** | **Membership** |

Fig 2: Extract from a HyperNews session

Feedback and Analysis

The Computers in Teaching and Learning component of MARBLE has experimented with some of the most innovative web technologies in pursuit of its educational aims. It has been crucial therefore to evaluate the effectiveness or otherwise of these methods in order that any necessary refinements can be identified to the course and that conclusions may be drawn relevant to the future of distributed computer-based learning. To this end, a detailed questionnaire was aimed at both student and tutor participants. The response rate for each category was almost 60 percent.

We had two means of assessing the popularity of HyperNews, the web-based many-to-many discussion forum: one was by the students' responses to the questionnaire, the other was by the access data collected by our server. All students kept in touch with HyperNews at least once a week, with most accessing it two or three times a week. Whether HyperNews was a good example of a peer tutoring tool is a more interesting question, however, and more difficult to assess. The students collective response was that they learned a great deal from HyperNews, though they were unable to be too specific about what it was exactly that they learned. Sixty percent of the student responses said that the HyperNews dialogue did help them learn in a general sense by giving them background that was not available through the coursenotes. It certainly demonstrated to them what their peers were doing and thinking and provided powerful motivation for them to take a more active part in the course. Some typical responses:

"Other people were thinking and it prompted me to think more about the stuff. I was more involved in the course because of the discussion."

"Probably did [help] generally, but not in specific answers for passing the exam. Made me think and focused my ideas."

Evidently, the underlying stated aim for HyperNews was that it stimulated dialogue between peers on issues that arose throughout the course. In response to this stated aim, the general feeling expressed by the students was that while HyperNews may not have directly supported the goals of the course, it did create a sort of learning community which provided motivation and possibilities for reflection.

The Answer Web, as described above, was another experimental feature of the CTL on-line course. Not enough students were aware of the distinction between the "ephemeral" nature of the discussions on HyperNews and the more "permanent" nature of the issues raised on the Answer Web, which functioned more as a tailor-made and infinitely expandable FAQ database. The feedback we received both from students and staff indicated that the Web was not quite ready for full-scale propagation: it needed to be "seeded" with more information and the students needed more help in understanding how to use this feature. One problem which arose, and to which we have no ready answer, concerns the comment of one staff member that he simply did not have enough time to respond adequately to questions directed to him through the Answer Web. Clearly, there is more labour involved in stocking and tending the Web in its early days and a more mature set-up should eventually reap benefits in a reduced workload eventually. We feel the Answer Web has exhibited interesting potential and the feature will be revised, re-stocked, and well-publicised to students participating in subsequent course sessions.

Conclusion

The MARBLE Project as a whole has managed in a very short time (7 months) to deliver on all of its 10 subprojects. Many innovative web-based elements have been built into each subproject. It is difficult to itemise succinctly those elements which have worked most successfully due to the disparate nature of the subject areas and techniques involved. However, useful conclusions may be drawn from the methods implemented for the CTL course, considered in some detail above. Here the emphasis for effective courseware distribution and design has been with a task-based, communication-led approach, rather than with content. The Web, through the mechanisms described above, is ideal for many-to-many dialogue between students and/or tutors; it creates a virtual forum which facilitates learning, discussion and reflection; it provides motivation and facilitates a fresh way of looking at things. It enables the presentation of an interactive task-based course structure which encourages students to think deeply about what it is that they know, and how to articulate it. The course structure is a reusable electronic architecture that can be adapted and utilised in a wide range of academic areas. If presented on the Web, the course *content*, on the other hand, is perceived by students to be in a less desirable form than on sheaves of paper. (This conclusion is not necessarily valid, however, for all the MARBLE subprojects. The *Geotechnical Engineering* subproject, for example, has successfully published material with a highly complex and graphical content; this material has found an effective medium in the Web where students may learn easily and interactively from the on-line maps and images.)

Rapid but effective solutions need to be found to the pressures which threaten Higher Education to bursting point with vastly increased student numbers and an expectation that graduate quality will be maintained. MARBLE is an experiment in addressing the three strands of effective educational courseware design: that resources should be reusable, that effort should be collaborative, and that teaching should be effective - and fun.

References

[Ackerman & Malone, 1990] M.S. Ackerman and T.W. Malone (1990). Answer Garden: a Tool for Growing Organizational Memory. Proceedings of the Conference on Office Information Systems '90. ACM Press.

[Duneen, 1996] Finbar Duneen (1996). Summary of students' responses to the role of network tools in the Computers in Teaching and Learning course", MSc Project Paper, Heriot Watt University

[Fowler et al, 1996] C.Fowler, J.T.Mayes, and B.Bowles (1996). Education for Changing Times. British Telecommunications Engineering, vol 15, part I, 1996, pp32-37.

[Laurillard, 1993] Diana Laurillard (1993). Rethinking University Teaching. Routledge

[MacFarlane, 1992] The Working Party of the Committee of Scottish University Principals (1992). Teaching and Learning in an Expanding Higher Education System. Polton House Press.

[Smeaton & Neilson, 1995] Calum Smeaton and Irene Neilson (1995). The Answer Web. Poster presented to the Fourth International Conference of the WWW, Boston. Available from http://www.orbital.co.uk/orbital/www5.html

Information on the conferencing system *HyperNews* is available at:http://union.ncsa.uiuc.edu:80/HyperNews/get/hypernews.html

The INTERACT Project is at http://www.icbl.hw.ac.uk/projects/INTERACT.html

MARBLE is at http://www.marble.ac.uk/marble/

Open Net and the Australian On-Line Education Market

Dr. Clark N. Quinn
Open Net Pty Ltd
Australia
info@opennet.net.au

Open Net is part of the Open Learning Initiative of the Australian Government. This initiative provides seed funding for three years after which the individual components are to be self-funding. Established in 1995 as an outgrowth of an earlier project in 1994, Open Net was set up to serve as an Internet Service Provider for the education sector.

Open Net established the only nation-wide local call Internet service in Australia. However, the telecommunications fee structure in Australia was such that Open Net could not continue the service and was redirected to the development and distribution of on-line educational content.

In Australia, as in most countries, education at all levels is heavily subsidised, and competing with these institutions is difficult. However, the institutions also have unrealistic expectations both about their own capabilities and about the financial returns possible. Thus, cooperating with these institutions is difficult.

To compound matters, the on-line market is still limited. While Australia has one of the highest per capita Internet connection rates, the overall population of Australia is small. Of course, the Internet is not geographically restricted, and the whole world is a potential market. The proximity to Asia and its burgeoning education market is an obvious connection, but that is not a new idea. Consequently, there is considerable competition.

In addition, the setting up of a web server has become a relatively straight-forward task, and many educational institutions have their 'explorer' groups that are transferring their learning materials to World Wide Web (WWW) pages and creating electronic mail support groups.

The above contraints have driven the business strategies for Open Net. We need to focus on a solid core of highly desirable offerings that would not obviously compete with existing offerings. We needed to differentiate ourselves on the basis of quality of product, flexibility of delivery, and the comprehensiveness of the solution. And, we need to make a profit.

Our strategy is, in short, to target our offerings carefully, source accredited offerings in this area, support conversion to on-line delivery, source the technological infrastructure, create a complete but lean administration, and market in ways appropriate to the medium.

In the course of our planning, we have conducted a thorough survey of the software packages that can support this enterprise, explored the on-line transaction capabilities available in Australia, contacted content providers around the world, and assessed the technological environment that suits the education market.

In this presentation, Open Net's current solutions to the problem are presented. As this is a moving target, what will be shown is a window on the environment as it has existed over the past year.

Procedures for Creating Useful Web Sites

Judy Cossel Rice
Department of Instructional Science, College of Education
Brigham Young University, Provo, Utah
Judy_Rice@byu.edu

Paul F. Merrill
Department of Instructional Science, College of Education
Brigham Young University, Provo, Utah
Paul_Merrill@byu.edu

Carol Lee Hawkins
College of Education
Brigham Young University Provo, Utah
Carol_Lee_Hawkins@byu.edu

Abstract. The Instructional Science Department at Brigham Young University developed a procedure for identifying the specific information that should be included at a Web site and an approach for addressing style and usability issues. The procedures focus on identifying the audience, determining their informational needs and designing a site where this information can be easily accessed. The procedure outlines the steps of preliminary planning, design planning (audience identification, audience needs assessment, content determination, cognitive design considerations, visual design considerations, usability facilitation) and Web site management.

Introduction

In the summer of 1994 we undertook the creation of a World Wide Web site for the Department of Instructional Science, Brigham Young University (http://www.byu.edu/insci). The site was to be a repository of information about our program and to be a reference for those enrolled in it. The project resulted in the development of a procedure for identifying the specific information that should be included at a site as well as methods for addressing style and usability issues. We combined task-centered and user-centered design principles to provide the theoretical foundation for the project. We are now using this approach to design a site for the College of Education.

Task-centered design principles focus on the tasks or jobs a targeted group needs to accomplish and what is necessary to complete the tasks. This approach facilitates the creation of a World Wide Web site by defining the information needs of a targeted audience. It requires designers to identify who will be accessing their site, what this group will be trying to accomplish, and what information they will need to accomplish the goal instead of simply placing a large body of available information online. Designers can then select what they wish to provide given the unique qualities of the Web and their resources.

User design principles focus on how people interface with programs--the human-computer interactions that occur. This approach encourages the designer to create a site where people can readily locate and retrieve the needed information. The designer pays particular attention to the graphical user interface (GUI), screen layout and access speed. People do not want to wait for several minutes while a huge image map loads, even if it is an extremely informative, graphically pleasing, navigational marvel. Similarly, if the user is unable to locate information at the site, it essentially does not exist for that person.

The Web was such a new medium when we first undertook the creation of the Instructional Science's site, we could not rely on established procedures for its development as none existed. While we were able to draw from other disciplines, we were still faced with issues unique to the Web such as limited formatting options, access speeds, download times, changing browser capabilities, standards and coding. The design sequence and features we developed evolved through may errors and repeated formative evaluations by users and focus groups as we struggled with answering the following six questions: What is the scope of our project? What content are we going to provide? How will we provide it? What will it look like? How will people locate the information? How will we manage the site?

Our experience with the Instructional Science site has allowed us to generate a guideline for the creation of future Web projects at the College of Education. New projects will go through the following development stages:

1. preliminary planning
2. design planning: audience identification, audience needs assessment, content determination, cognitive design considerations, visual design considerations, usability facilitation
3. management planning

We will share these guidelines with you as well as some of the mistakes we made, the difficulties we encountered, and the lessons we learned during this project. We will first present the guideline and then discuss how or why it evolved.

Procedures

Preliminary Planning

The designer begins by meeting with the person commissioning the project and determining the general goals, degree of autonomy, time frame, funding, resources (both people and equipment), and evaluation procedures. A team is formed and a project management book is started with goals, due dates, and person responsible. Initially, we had one person assigned to do the entire project but we discovered it became an impossible task as Web authoring became more complex. We found it necessary to establish a team with members specializing in graphics, Java Script, multimedia, human-computer interaction and instructional design in response to new browser capabilities. The team concept also makes it easier to adhere to the project goals and time frame. When only one person authored the pages, it was easy to perseverate and spend an inordinate amount of time perfecting and exploring the potentials of design features. The design team can work together to solve problems and make sure deadlines are met.

Design Planning

Once the parameters are defined, the next step is to determine the content to be included on the site. With the inclusion of hyperlinks, the Web allows for the creation of a site with access to almost limitless information. The design challenge is selecting relevant material. This selection process becomes easier once you identify the audience and establish their information needs. While this process may initially seem time-consuming, it facilitates the creation of a quality site.

Identify the audience. The project sponsors decides who they want to encourage to access their site. Our department chairman chose to focus on prospective, new and continuing students for the Instructional Science site. The purpose in identifying the audience is to allow the design team to concentrate on the needs of a particular group. Without such a focus, the site development effort may wander and may lack coherence.

Determine the audience information needs. After the design team knows who the audience will be, the next step is for them to determine the audience's information needs. This involves four steps:

1. select a few members of each group within the target audience
2. interview them and conduct a task analysis
3. generate a work model
4. do a role delineation

One can get a general idea of the of the needs of a group by interviewing one or two individuals from the group. We selected a perspective graduate student, two new students, and two continuing students to represent the three groups in our targeted audience. Their responses were adequate for us to get an idea of what they needed. This is where a team is beneficial--each member can interview a few people and the job is not too time consuming.

When a team member conducts the interview, have the individual describe their job and what they do each day. For example if they are students, have them talk about how they selected their classes, enrolled, how they go about completing assignments, study, etc. Ask what resources they use. Record their responses (a tape recorder is helpful). After the interview, the team will generate a work model (depiction of the task analysis) by organizing these responses into a structured list or diagram. Often the person will skip from topic to topic during the interview and restructuring their responses allows the interviewer to make sure they understood the person correctly and to determine if the information is complete. Show the completed work model to the person interviewed for review and revisions. The team next combines the separate work models into one. The process is repeated for each identified group and separate work models are created.

The completed work models for the different groups provide valuable data on the tasks each person performs and their information needs, but are not very helpful in seeing commonalities between groups. To do this, the team will need to do a role delineation or restructuring of the data. Each of us assume many different roles during the course of a day i.e. learner, teacher, parent, spouse, cook, planner, decision maker etc. Identifying these roles allows the team to compare the information needs of different groups assuming the same role. While various groups may use information to perform different tasks, they often need the same information as they assume the same role. For example, a secretary may need to call a professor to request information on a proposed addition to the catalog. A professor may need to call a colleague to ask they substitute for them while the professor attends a conference. A student may need to contact a secretary to request a study list change. While each task is different, each person is assuming the same role of communicator and each needs the same information: the telephone number of an individual.

To generate a role delineation the team gathers the work models for a particular group and sees if they can assign general categories describing the type of task being performed. While the task analysis breaks the activities down into small steps, the role delineation reassembles the steps in groups of similar actions. As we looked at the graduate student work models, we were able to break down the tasks into making decisions, planning, finding out about correct procedures, doing assignments and communicating with others. These categories were then used to identify the roles of decision maker, planner, procedures expert, product producer, and communicator. The information contained on our site would need to enable the student to fulfill these various roles.

The power of the role delineations is that one can look at them and see where the roles and the information needs overlap. The more the overlap, the greater the need for that piece of information. This higher need allows the designers to set priorities for sequence of page development. A single page can be designed which will satisfy the needs of all assuming the same role. For example when we saw the role of communicator emerge in each group we interviewed, we knew there was a strong need for information to facilitate the tasks a person acting in this role performed. There was already a telephone directory provided to students, but often those interviewed expressed that they needed to call the operator to get a number because the directory was not at hand when it was needed. Additionally, the students did not have access to the directory listing email addresses for faculty--this information could only be obtained from the department secretaries or the professor. Potential students and alumni from off-campus were only able to obtain telephone numbers, not addresses. There was no single directory readily available for the different groups to provide the needed information. This commonality allowed us to determine that there was a great need to provide a way of contacting people. Each role was similarly examined and the content needs and priorities established.

Determine content. Decide which needs may best be met given the unique qualities of the Web. Just because one may put something on the Web, doesn't mean it should be there. Each medium has its strengths and limitations. Sometimes face-to-face conversation is the best means of communications.

Other times printed materials, videos, or CDÕs may be the superior delivery method. In order to determine which materials should be delivered via the World Wide, we needed to identify how it was effectively being used in education and what its capabilities were. When we looked at various academic sites we saw six functions emerging: identifying oneself, providing information or instruction, creating a research environment, furnishing a medium for publishing, encouraging discussion, and establishing a platform for consultation and collaboration. While the Web was being used for other functions (humor, commercial, etc.) we wanted to restrict our usage to providing materials of worth for the community of learners and educators. Next we identified the WebÕs capabilities for data display to determine how these features could best be utilized in our design. The Web provides hypertext links, multimedia, and interactivity. Therefore the material we were to provide needed to fulfill the needs of our audience, fall into one of the six functions, and take advantage of one of the capabilities of the Web in order to be justified as being included on our site.

In the case of the role of communicator discussed above, the creation of a central database of how to contact people in Instructional Science would be a justifiable endeavor (an information function). We designed pages listing each faculty, staff and student with office addresses, telephone numbers and email addresses. We wanted to avoid simply creating an electronic directory which could be better provided in print form. Instead we examined the features of the medium and created a page where the users could not only locate telephone numbers and addresses but contact the party directly from their browser using the "mail to" feature in Netscape. Our work models also indicated that frequently people were unsure of who could help them so calls were made trying to identify the faculty or staff best able to answer their questions. Once the party was identified, additional problems were encountered trying to make contact with the individual as frequently the party was not in their office. Thus our pages were designed with hyperlinks for each person listed to provide descriptions of their duties, interests and office hours to facilitate making contact with the correct individual utilizing the unique hypertextual feature of the Web.

At the completion of this phase of development the team will have the work models, role delineations, list of information needs and the proposed site content so that they may turn their attention to the design issues.

Design Process

Cognitive design considerations. Cognitive design refers to creating documents which address how the brain receives, organizes, stores, recalls and utilizes data. Good design facilitates the integration of data into existing schema by placing it in the context of established knowledge. If you want users to know how to locate information at your site, organize it in a logical, predictable manner. Good metaphors can simplify this integration by establishing parallels: reference documents can be located in a “virtual library”; chat forums can be located in the “faculty lounge”.

The brain can deal with information easier if it is sequenced and grouped into logical units. A long list of topics can lead to cognitive overload whereas a list organized into seven to nine natural categories is more usable. One of the more challenging tasks we initially encountered in designing documents was how to create whitespace to offset and highlight this grouping of data. While lists, buttons and icons helped, it was the introduction of tables and the ability to color the background of individual cells that have allowed us to control layout sufficiently to create truly user-friendly documents.

People have a need to see the logic in the organization of a Web site just as they do for any data. Presenting an overview of the site’s structure provides an orientation mechanism just as a map does for an unfamiliar terrain. People feel comfortable when they know where they are and can anticipate where they need to go to find something.

We were able to use the roles we identified in the role delineation as a foundation for organizing the material at our site.

Visual design and layout considerations. The design team next decides how the information will be presented. Visual design addresses how pages will look and how information will be displayed. The Web was such a new medium when we undertook the construction of our Web site that not many guidelines existed. We endeavored to develop style guidelines of our own by surveying existing sites and analyzing what seemed to work for us visually and what did not. Additionally, many of the principles of screen design for hypertext and electronic documents were applicable for Web pages. As we surveyed different sites on the Web, we found we were unwilling to read text-intensive pages and preferred those with concise, bulleted list. We concluded that the sites we judged as good were those based on consistency (information structured in predictable ways), simplicity (no gratuitous graphics or Óeye candyÓ), redundancy (using the same backgrounds, colors, graphics, layout on similar document) and clarity (concise, well organized material).

Since reading text on a computer screen is generally slower than text on a printed page, it is particularly important that the design team carefully considers how much text will be on a page and how it is organized if they want users to read them. There should be a balance between text and graphics. Graphics should be chosen on the bases that they are indeed the best way to communicate a concept. Pages should be kept short (1 1/2 to 2 screens) whenever possible. A carefully organized design that is consistent across pages will aid users in quickly finding the information they want. Good design also adds credibility to the site as it reflects the professionalism of the designers.

We reformatted much of the text contained in documents that couldn't be reduced to one screen. Students had complained they had been unable to easily locate a particular section in documents when they had to scroll. A hyperlinked menu was provided at the beginning of those longer documents to aid in accessing the subsections. Bold type, headings, subheadings and lists were used in documents to facilitate locating specific information and topics. We created a standardized format for pages containing the same information, i.e., vitae, class descriptions, labs, resources and syllabi, so that users could more easily locate the desired information. This redundancy enabled users to anticipate where to look on the page. Subsections were divided by a colorful line again for ease of quickly scrolling through material.

Usability Facilitation The graphical user interface (GUI) or home page is the most important page the design team will create as it contains an overview or index of the materials at the site. It sets the tone for the site and provides the means of navigating the hyperlinked documents. The index on the home page should not be too shallow (containing too many topics) nor too deep (containing too many subtopics). A shallow index results in a long list of materials and one that is too deep list only a few items and leads the user to yet more index pages. The ideal page should be a well balanced hierarchical tree that affords quick access to information and helps users understand how things are organized. The material on our Instructional Science home page was divided into ten major categories of information each with three to four subcategories. Our goal was to allow the user to access any information with three links.

The categories and icons were chosen after repeated user testing. We first attempted to create the categories of information on our own by reviewing the documents on our site. Students expressed confusion in locating information so we attempted further testing to refine the divisions. Students were asked to sort strips of paper with the titles of separate documents into categories of their choosing. The categories were compared to the ones we had developed and a revised page was designed. The students were then asked to predict where they would find information given the ten categories on the new page. The success rate was lower than desired, so subcategories were displayed to facilitate locating data. The revised page containing the categories and subcategories was then retested resulting in a much higher success rate.

Once the layout of the home page was resolved, the next step was selecting icons for the categories. One user test involved students matching icons to the categories they thought they represented. Another test had the icon and the student was asked to state a category it suggested. After going through numerous icons and revision, the final selections were incorporated into the design of our home page.

Our current home page has a brief graphical menu for browsing and a separate comprehensive listing of all the documents organized by categories for more direct access to information. The background pattern was selected for simplicity, aesthetic appeal and to serve as a signature to let the viewers know when they were accessing materials on our server and when they had linked to another site.

Additionally, and of equal importance, are usability factors of ease and speed of use. Studies on user response to computing system delays suggest that people are unwilling to waiting longer than about 20 seconds. While large image maps are dramatic, they normally take longer to load, frequently far exceeding the 20 second ceiling of viewer tolerance. We have chosen to minimize the number of graphics we use until modem speeds dramatically increase.

Management Planning

It is important to establish procedures for managing a Web site from its inception. How to organize materials on the server, how to keep the links updated and when to revise the site need to be clarified. Before any work can begin the designer needs to establish the document organization structure for materials that will be placed on the server. He needs to ensure that this structure has logical breadth (top level directories) and depth (embedded subdirectories) before any documents are authored because all links will depend on this directory/subdirectory arrangement. All names should be simple and in lower case. If a user is trying to access a particular document by keying in its URL, the directory and document names should be easy and short. We had not anticipated the ramification of naming directories and deeply embedding documents. Trying to correct many of our early errors has become an extremely time-intensive job. We have had to make several revisions to all our documents due to poorly designed directory structure, the use of upper and lower case names and a server change. When we had our own server the deep directories had not seemed a problem and the use of upper and lower cases seemed to make it easier to read the document name. However when our site was a moved to a university server, we had document names such as: http://www.byu.edu/acd1/ed/InSci/People/Faculty/MerrillVita.html. The need for careful naming is obvious the first time one tries to say that URL aloud specifying the correct cases.

Off-site links need to be routinely checked to see if they are still current. We have created numerous directories of information for our students. It generally takes about an hour to validate all the links on one of these documents and to update it with new additions. Scheduling time for this routine maintenance is vital. If users come to a site and many of its links are useless, they will stop coming. Not only do links need to be updated, the general site design also needs review as new features become available through extensions to the HTML code. Keeping current with these expanding capabilities of browsers has become one of the major challenges for us. The technology is evolving so rapidly that as new features become available, new solutions to navigation and layout problems become feasible. While creating added difficulties for us designers, (we are continually revising pages) these features have allowed our site to become increasing more user friendly.

Conclusion

One of the major lessons we learned in developing our site is that the design process is not a linear, purely sequential process. It was more akin to laying track under a moving train as new features were introduced. We created a document, tested it, worked on resolving its flaws, revised it just in time for new features to be introduced opening up possibilities for better solutions. One major challenge we faced was knowing when to settle for what we had. We are still fine tuning the documents and will be shortly completely revising the site in response to the lessons we learned and giving the site an updated look.

Another major lesson we learned was the importance for continual formative and summative evaluations using different browsers and different platforms. We would design a document we felt was very readable and would put in what we felt was a logical location. When we asked a student to locate it, or even more embarrassingly, when we would try to locate it later, it would have seemed to disappear. Its location was no longer obvious. User testing revealed several such weaknesses. Evaluation also alerted us to problems with slow load time and platform specific difficulties. We developed our documents on Macs using the latest versions of Netscape. Sometimes when we viewed them on other platforms, or other browsers, the document were incredibly ugly or unreadable.

Our experiences have motivated us to make changes in how we create Web projects in the College of Education. We have created policies and guidelines for the College of Education Web use. We have now established a Web Advisory Team to oversee and to provide direction for our projects. We have formed a Web Development Team utilizing both undergraduate and graduate students interns. We are now devoting part of the team's development time to inservice where we can keep current with the rapidly evolving technology. We have assigned a student to maintain our resource documents.

The creation of useful Web sites takes a commitment of time and resources. It takes effort and careful planning. The results of these efforts, however, are sites that satisfy the needs of those accessing them.

Examples and references may be seen at:

http://www.byu.edu/acd1/ed/workshops/design.html

Professional Electronic Publishing in Hyper-G

The Next Generation Publishing Solution on the Web

Klaus Schmaranz
(Graz University of Technology, Austria
kschmar@iicm.tu-graz.ac.at)

Abstract:

The first part of the paper identifies disadvantages of first generation Web publishing solutions that have to be overcome for professional publication providers. Using Hyper-G for distribution of electronic documents opens the way to the first fully integrated professional publishing solution on the Web. User and group access rights as well as billing mechanisms are integrated in the server, links are bidirectional and annotations to existing documents are possible without changing the documents' contents. Hyperlinks are supported in arbitrary document types besides hypertext which opens the way to new publishing paradigms beyond the paper based approach. Naturally a rich set of structure and search mechanisms is provided. On this basis a set of tools has been developed that supports electronic refereeing, automatic hyperlink creation, glossary links and table of contents generation nearly fully automatically. All the data prepared on a Hyper-G server can then simply be exported to CD-ROM without any additional effort which allows hybrid Web/CD-ROM publications.

1 Motivation

Electronic publishing is today one of the booming branches on the Web. Web surfers will find lots of electronic paperware on their ride through the servers. However most of the Web sites that serve electronic publications are run by universities and only a few are operated by publishing companies on an evaluation basis free of charge.

Having a closer look at the way electronic publishing is done today one will mostly find HTML or PDF documents that are very similar to their paper based counterparts. Often a search engine is provided to make location of interesting papers easier, all other benefits of doing publishing electronically are mostly neglected. For the reader of electronic publications nearly no value is added compared to paper based articles. Worse than that - considering HTML documents the possibility to do high quality printouts for archival purposes is lost. This is surely not enough to make electronic publishing on the Web a success.

Distributing electronic documents on the Web is considered to be cheaper than publishing on paper [see Odlyzko 95], additionally turnaround time from submission of an article to the final published version is considered to be significantly shorter than for paper based publications. As experience with J.UCS [see Section 5] shows, this is certainly not true for high quality publications that are refereed. In this case refereeing, corrections by the authors and resubmission are the most time consuming parts of document preparation. Once the paper is ready for publication hyperlinks have to be inserted in the electronic version, which is in first generation systems still done by hand because appropriate tools are missing. Both refereeing and hyperlink creation can be speeded up when using the right tools as can be seen later.

Hyperlinks in first generation Web systems are represented by URLs (Uniform Resource Locators) [see Berners-Lee 94] that are unidirectional and embedded in documents. This approach implies that links can easily point to nowhere (the ``this document has disappeared syndrome"). Using URNs (Uniform Resource Names) instead of URLs is the solution to this problem. URNs do not longer point to the physical location of a document on a server but add another level of indirection by defining unique location independent names for documents.

Embedding hyperlinks in documents has the disadvantage that only document formats designed to be hyperlinkable allow navigation through cyberspace as is the case with HTML. Unfortunately a lot of formats used for different document types were designed long before anybody knew about the Web. This includes all image formats like TIFF, GIF, JPEG, [see Murray 94], it also includes all video formats such as MPEG as well as one of the most widespread formats for professional publishing: PostScript [see Adobe 90]. The way out of this dilemma is to have a separate link database and consider links to be overlays. This automatically makes all document formats hyperlinkable because documents and links are handled separately.

Structuring and classification of documents in standard first generation Web servers has to be hand-coded using HTML documents. For this reason poor structure and no document classification is provided on most servers because it is too much work. This makes navigation rather difficult [see Andrews 95], because after a having followed a few hyperlinks the readers no longer know where they are (the ``lost in hyperspace" syndrome). Once a reader has found an interesting paper it is absolutely necessary to immediately store a bookmark. If the way to the paper was longer than e.g. 3 mouseclicks it is nearly impossible to remember the way the next time it is needed.

One of the really severe problems of electronic publishing today is the lack of accounting and billing mechanisms as well as user access rights [see Maurer 94]. For this reason electronic publications on the Web are mostly free of charge which is not in the sense of publishing companies.

Very often it would be interesting for server operators to have statistical information on a user session basis to be able to follow users on their way through a server and optimize their offer. As an example one could find out that 60% of the readers from overseas are leaving the server after having followed 4 hyperlinks. Having a closer look it could come out that the document after the 4th hyperlink contains a huge inline image that is inacceptable for an overseas data transmission and that this document has to be changed.

2 Hyper-G as a Basis for Electronic Publishing

All the problems and difficulties mentioned above have lead to the development of a Web server with a completely new underlying concept compared to first generation Web systems: Hyper-G [Maurer 96]. Hyper-G now allows it to really do professional electronic publishing on the Web and to enter a new era of electronic publications in terms of usability and content.

Up to now electronic publishing was either done on the Web or on CD-ROM basis. Using Hyper-G one can do hybrid Web and CD-ROM publication without additional effort as has been successfully proven over the last two years with J.UCS, the Journal of Universal Computer Science by Springer Pub. Co. [see Maurer 94]. The only thing publishers have to do to produce a CD-ROM is to prepare the data on a Hyper-G server and then export the whole collection tree or parts of it.

The kernel of the Hyper-G server is an object-oriented distributed network database with a separate link database. Information structure as well as document meta information are a basic part of the concept [see Kappe 91]. This makes it possible to present the user a seamless world-wide structured information space across server boundaries.

Document meta information such as author, title, keywords, creation date, modification date as well as expiry date and many more support the Web surfer in getting as much information as possible. Naturally document meta information is searchable as well as fulltext search can be performed. The scope of searches is user definable and can be one small part of one server or even the whole content of all servers worldwide in one single operation. Even when doing searches on multiple servers it is not necessary to know about the server addresses.

More than that: meta information cannot only be applied to documents but also to hyperlinks! This means that links can have types, such as annotation links, inline links, also version links for documents where multiple versions exist and many more.

Hyper-G servers do not only provide read access but also write access is possible. Read and write access to documents are controlled on a user and group access right basis and billing is integrated in the server.

This concept opens completely new perspectives of electronic publishing: having the hyperlinks in a separate link database makes every document hyperlinkable even when the document format does not allow links [see Maurer 96]. All links in Hyper-G are bidirectional, making it possible to not only follow the links pointing from one document to another but also to see links referencing a document and follow them in reverse direction. Being able to examine the neighbourhood of a paper makes it possible to find other interesting papers on the same topic that very likely are difficult if not impossible to locate if only unidirectional links were possible as is the case with first generation Web servers.

Hyperlinks in arbitrary document types such as PostScript, images, movies, 3D scenes and even sound make navigation in hyperspace easy and a structured hierarchical view of the database with location feedback helps overcoming the ``lost in hyperspace" syndrom.

With this approach electronic publications need no longer be text based with some multimedia add-ons. Instead authors can choose the document type most suitable for the topic without loosing important hypernavigation features.

As an example a paper about new chemical structures could consist of 3D models of molecules that are clickable. The hyperlinks could then lead to spectrum images that are then linked to some additional text based explanations in e.g. PDF [see Adobe 93]. A video of an experiment, naturally again with hyperlinks to explanations, completes the presentation.

All the documents in the example above carry meta information such as keywords and can therefore easily be located in a search.

Acceptance of electronic publications is highly dependent on their quality. For electronic publishing quality does not only mean high quality contents, which can be assured by an appropriate refereeing process. Stability of electronic publications is at least as important. Technically it is easy to change electronic papers after publication but this is unacceptable. Instead Hyper-G's annotation and versioning mechanisms can be used to alert the reader of new results or errata. In this case the paper is not changed at all, only additional information is added to the paper. Therefore all citations of the paper that existed for the original version are still valid and the reader can choose to browse annotations and newer versions of the paper on demand.

Annotations in Hyper-G are hyperlinks pointing to the document that is annotated. Since Hyper-G's links are bidirectional the reader simply follows an annotation link backwards to read the annotation. The use of URNs in a link database instead of URLs embedded in documents guarantees that the annotation links are stable. This means that an annotated document can be moved around on the server or even from one server to another without generating annotations that point to nowhere. All links that pointed to the document before are then pointing to the document at its new location.

As has been mentioned above refereeing is one of the very time consuming processes of publishing. Using annotations electronic refereeing can be performed easily: papers are inserted in the Hyper-G server with read access only for the referees. The referees then comment the papers using the annotation mechanism. If desired annotations can also be made readable for the author, so the author is able to react immediately on the referees' comments. More than that - the author himself could also annotate the referees' comments to clarify misunderstandings. Naturally the author as well as the referees remain anonymous [Maurer 95]. Saving the time for sending documents back and forth between referees, authors and editors as well as being able to do corrections in papers while refereeing is still in progress this process can be shortened significantly.

Amongst other structuring elements Hyper-G supports the concept of clusters. A cluster contains several documents that are related to each other and therefore should be viewed together. In the example given earlier in this paper the 3D molecular models could be clustered together with an explanatory text. In this case the user would get the 3D model in one window together with the explanatory text in another window.

Clusters are also used to serve multilingual documents. Documents in different languages are put together and the readers then only get the documents matching their language preferences. In first generation Web systems the only possibility to have multilingual documents was it to let the user choose the language on the entry page and then follow different paths through the server for different languages. This approach caused a lot of work for server operators and the readers had no chance to change the language on the fly. With Hyper-G only one path through the server has to be maintained and the readers can switch back and forth between multiple languages on the fly.

3 Professional Tools for Publishers

So far discussion was about the technical design of Hyper-G and the resulting possibilities arising when using it for distribution of electronic documents. But that's not really all: these features can also be helpful for internal purposes such as collecting all the versions of a paper from the original submission over intermediate corrected versions to the final published paper. The complete refereeing process including all different versions of papers together with the referees' comments can be kept on the same server that is used for distribution. Supported by Hyper-G's access control mechanisms the publisher defines what a reader can see. As an example subscribers can only see the published versions of papers, referees would also see intermediate versions of the paper they are just refereeing. The editor in chief and the responsible staff would have access to all the information including refereeing forms and internal information. Referees would not have the rights to change a paper they are refereeing, but they would have the rights to make annotations. If desired the author could get the right to also annotate the referees' annotations to clarify misunderstandings - naturally both referees and author remaining anonymous.

For the whole process from administration to final document preparation a set of tools and fill-out forms has been developed that can be adopted according to the publishers' needs. The following paragraphs deal with some of the more sophisticated tools that have been developed to provide a cost effective way for publishers to add value to electronic papers.

As mentioned above it is desirable to keep older versions of documents for archival and maintenance purposes. For this reason a special versioning tool was developed that uses Hyper-G's cluster mechanism: different versions of a document are clustered together and the reader can switch back and forth between different versions on the fly. Whenever a new version of a document is inserted a special link migration tool parses all outgoing hyperlinks in the older document and inserts them at the appropriate location in the new document. Only outgoing links are considered during migration because incoming links should not be touched - they are normally comments for exactly the version of the paper they are pointing to.

Considering a paper one will find a lot of so-called vocative links. A vocative link is a textual pointer to a location such as ``see also page nn". In scientific papers one will normally find a references section with pointers to other publications - again a typical example for vocative links.

These considerations about vocative links have lead to the development of an automatic vocative link creation tool. The tool is based on a so called ``Vocative Link Creation Language" (VLCL) that was designed to support the description of contexts in documents and to find potential hyperlinks in it. To not loose myself in too many details, here is a small example:

Consider a piece of text with the phrase *...details can be found in [Moser 95]....* The tool will identify this phrase to be a vocative hyperlink leading to the references section and insert a link to the references here. Parsing the references it finds the entry *[Moser 95] Moser J., ``The Art of PostScript programming", available at http://www.iicm.edu/app, Dec. 95*. The URL will be identified by the tool and a link to the according document will be inserted. Since VLCL is a programming language the behaviour can even be controlled to the extent that the link to the references is not created but leads directly to the location instead.

If this sounds complicated - VLCL was designed to describe such rules in a very compact and high level manner. Normally VLCL programs having the functionality described above are not longer than e.g. 30 -- 40 lines. And they normally have to be written only once because journals have their well defined citation rules that are not changing too much over time. In addition citation rules are not varying too much between journals so that an existing VLCL program can be slightly modified and will then suit the needs of another journal.

A different tool for automatic hyperlink creation deals with glossaries. A glossary in Hyper-G is defined as an arbitrary collection of explanatory documents that are classified by their titles and keywords. The glossary link creation tool accepts arbitrary many glossary collections and automatically interlinks items in papers with glossary entries. The links created get the special type *glossary* so that the reader can turn them on and off seperately when needed. The glossary link creation tool works at the moment for HTML documents, PDF and PostScript support are under development.

Feedback about the readers' behaviour is necessary to improve the quality of a server. Standard statistics tools today are able to count the number of accesses to a document and give information about the location of the reader. Due to Hyper-G's user session concept a lot more information can be extracted from the logfiles. For this reason a specialized statistics tool was developed that is also able to trace the readers' way through the server from the beginning to the end of a session.

Using the session-oriented statistics the information provider can see the ``typical" way of a reader through the server and find out about specific problems that arise. As an example the server operator could see that a majority of users is quitting a session when downloading a certain document. In this case the information provider could have a look at this document and e.g. find out that there is a huge inline image on this page that takes to much time to be downloaded. Another situation could be that link references are misunderstood and the readers get a page they don't want.

It can also be analysed if the structure on the server is easy to handle for the reader. Items that are often searched instead of accessed directly are very likely not reachable easy enough. Parts of the structure that are never accessed also alert the operator that something must be wrong there. Analysing access and path statistics carefully and inserting appropriate links or restructuring the server accordingly helps a lot to ensure proper quality.

4 Making the Readers' Life Easier

Up to now discussion was about how to add as much information as possible to documents. One of the points was to interlink documents to the maximum extent possible by adding inter- and intra-document links as well as glossary hyperlinks and more. Using the tools described above this can be done with minimum effort and stability of hyperlinks is guaranteed by Hyper-G. But there is a negative aspect too: too many hyperlinks also means too many color changes which disturbs the reading flow.

As has been discussed earlier Hyper-G supports different kinds of hyperlinks: annotations, glossary links, referential hyperlinks, inline links, texture links and many more. To not have to many disturbing color changes links of different types can be turned off and on seperately when needed. Another possibility that is at the moment only implemented in Harmony is to show different kinds of links in different colors (Harmony is one of the very specialized Hyper-G authoring tools). Utilizing this feature the reader would know about the type of a link before even following it.

As has been mentioned before grouping of documents in clusters is one of the basic concepts in Hyper-G. One of the applications of clusters is it to provide multilingual versions of the same document. Naturally it is normally not possible to translate every single document on a server to several different languages. On the other hand it is not too much effort to translate neuralgic and mostly static parts of the data such as glossaries. In reality if there are multilingual glossaries existing it normally does not even help too much to translate the papers as well.

Consider an international journal: the language of papers will most likely be English and readers will understand English. The only problem is that some very topic-specific words will be found in the papers, in which case readers can use the glossary to look up unknown words. If the readers are not native English speakers and the glossary also contains a difficult explanation they would be forced to follow some more links between glossary entries until they finally understand the meaning. For this group of readers a translation of the first entry would very likely help a lot. And there is certainly a high number of non native English speakers reading English papers.

An all-time-important aspect on the Web is transmission speed. Long distance data transmissions on a slow transmission rate are annoying and in the worst case electronic publications will not be accepted by the reader for the lack of availability. There are two ways out of this dilemma that can be gone simultaneously: providing documents of different size and quality as well as mirroring and caching documents.

For providing documents of different size again Hyper-G's cluster concept can be used. A cluster could e.g. hold the same document in HTML format with small inline images for quick browsing, the better version could also be HTML but with high quality images and the professional version could be PostScript with 600 dpi resolution including 600dpi true color images. The reader decides once per session which level of quality to get and the rest is done automatically throughout the whole session. Naturally readers can change their choice on the fly.

More than that - consider the case of an HTML document with different sets of inline images - small ones with poor quality and huge professional ones. On standard first generation Web servers two different versions of the HTML document must be provided with different links. On Hyper-G the links are not directly pointing to the images but to the clusters! In this case the HTML document has to be prepared only once and the inline images automatically change according to the readers' choice, even on the fly if desired!

The other well known solution to overcome the transmission speed problem is it to cache and mirror documents. Naturally Hyper-G has the standard proxy functionality implemented. Besides that a sophisticated document replication mechanism [see Kappe 95] is implemented in Hyper-G.

Document replication in Hyper-G is understood to be a special kind of mirroring with the mirrored documents knowing about the original. This is possible by giving the documents a replica identification that matches the global object Id of the original document. The global object Id of documents in Hyper-G is a world wide unique 64 bit identifier.

The effect is that documents can be mirrored to the local site without the users having to know about. If they try to access one of the original documents on the remote server or even one of the documents on another mirror site they automatically get the replicated document from the local site instead of doing a long distance data transfer.

5 Current Electronic Publications With Hyper-G Technology

- The first electronic journal based on Hyper-G was J.UCS - the Journal of Universal Computer Science by Springer Pub. Co. It is a monthly journal covering all knowledge areas of computer science. Additionally to the Web version a yearly CD-ROM and printed version are provided by Springer. Papers in J.UCS appear in two parallel formats: hypertext and hyperlinked PostScript, PDF is planned for 97. See [http://www.iicm.edu/jucs] - the master J.UCS server - for more details and information about mirrors.
- Academic Press decided to distribute JNCA - the Journal for Network and Computer Applications (former JMCA and JMA) - electronically using Hyper-G starting in January 96. JNCA can be found at [http://www.iicm.edu/jnca], mirrors are in preparation.
- One of the most reputable journals in physics, FBS - Few Body Systems - also by Springer - started regular electronic service in January 1995 using Hyper-G. Find FBS at [http://www.iicm.edu/fbs]

- The German bible of Data structures - Datenstrukturen by Ottmann and Widmeyer - is published electronically on a Hyper-G server using hyperlinked PostScript. At the moment the german version is available at [http://www.iicm.edu/datenstrukturen]
- Since this paper excessively deals with Hyper-G - naturally also the ``Hyper-G bible'' called ``HyperWave - The Next Generation Web Solution'' published by Addison Wesley is available electronically at [http://www.iicm.edu/hgbook] (Note: HyperWave is the product name of the software whereas Hyper-G is the name for the underlying technology). Check out [http://www.iicm.edu/hyperwave] for detailed information about Hyper-G technology and the HyperWave product line.
- Additionally to the Hyper-G bible Addison Wesley decided to publish some 30 books electronically on the Web using Hyper-G. Some of them are already under preparation. As soon as they are released they can be found in the IICM electronic library [http://www.iicm.edu/electronic_library].
- One of the most comprehensive German encyclopedias - Meyer's Lexikon - is electronically available via Hyper-G on an n-user license basis - have a quick look at [http://www.iicm.edu/ref.m10] if there is something you always wanted to know.

References

[Adobe 90]
Adobe Systems Inc.: ``PostScript Language Reference Manual'', 2nd Edition, Addison-Wesley (1990)

[Adobe 93]
Adobe Systems Inc.: ``Portable Document Format Reference Manual'', Addison-Wesley (1993)

[Andrews 95]
Andrews K., Kappe F., Maurer H. and Schmaranz K.: ``On Second Generation Network Hypermedia Systems'', Proc. ED-MEDIA '95, (1995), 69-74.

[Berners-Lee et al 94]
Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. and Secred, A.: ``The World-Wide Web.'' Communications of the ACM 37,8 (1994), 76-82.

[Kappe 91]
Kappe F., Maurer H. and Tomek I.: ``Hyper-G - Specification of Requirements'', Proc. Conference on Intelligent Systems (CIS) '91, (1991), 257-272

[Kappe 95]
Kappe F.: ``A Scalable Architecture for Maintaining Referential Integrity in Distributed Information Systems'', J.UCS 1,2, (1995), 84-104.

[Maurer 94]
Maurer H. and Schmaranz K.: ``J.UCS - The Next Generation in Electronic Journal Publishing'', Computer Networks and ISDN Systems, Computer Networks for Research in Europe, Vol. 26 Suppl. 2, 3, (1994), 63-69.

[Maurer 95]
Maurer H. and Schmaranz K.: ``J.UCS and Extensions as Paradigm for Electronic Publishing.'', Proceedings DAGS'95, Boston Massachusetts, (1995).

[Maurer 96]

Maurer H. ed.: ``HyperWave - The Next Generation Web Solution", Addison-Wesley, (1996).

[Murray 94]
Murray J., D. and van Ryper W.: ``Encyclopedia of Graphics File Formats", O'Reilly and Associates, Inc. (1994).

[Odlyzko 95]
Odlyzko, A., M.: ``Tragic Loss or Good Riddance? The impending demise of traditional scholarly journals." In ``Electronic Publishing Confronts Academia: The Agenda for the Year 2000," Robin P. Peek and Gregory B. Newby, eds., MIT Press/ASIS monograph, MIT Press (1995).

WebGrid: Knowledge Elicitation and Modeling on the Web

Mildred L. G. Shaw and Brian R Gaines

Knowledge Science Institute
University of Calgary
Alberta, Canada T2N 1N4
{mildred, gaines}@cpsc.ucalgary.ca
http://ksi.cpsc.ucalgary.ca/KSI/

Abstract: This article reports on experience in moving a highly interactive application, that of repertory grid elicitation and modeling, from personal computers to World Wide Web servers to make it widely available on a platform-independent basis. Repertory grid elicitation is a knowledge acquisition technique developed in personal construct psychology that supports the development of conceptual models by eliciting the distinctions made by individuals construing a domain. It is widely used in clinical psychology, education, management and knowledge-based system development. This article illustrates repertory grid elicitation and modeling over the web in a personal decision making domain, that of selecting a home from those offered by a realtor.

- 1 Introduction
- 2 WebGrid in Action
    - 2.1 WebGrid Set Up
    - 2.2 WebGrid Triadic Elicitation Facilities
    - 2.3 WebGrid Other Elicitation Facilities
    - 2.4 WebGrid Modeling Facilities
- 3 Conclusions
- References

1 Introduction

Repertory grid elicitation is a knowledge acquisition technique developed in personal construct psychology (PCP) that supports the development of conceptual models by eliciting the distinctions made by individuals construing a domain (Gaines and Shaw, 1993a). The technique uses concrete instances of elements relevant to the domain to provide an easily conceptualized environment for elicitation, and prompts the client with a range of questions about similarities and differences in the domain that lead to rapid elicitation of the distinctions made, and the terms used for them. Repertory grids were originally developed to elicit the conceptual models of clients undergoing psychotherapy (Kelly, 1955), but were rapidly found to be useful in a wide variety of applications in management, education and the development of knowledge-based systems (Shaw, 1980; Shaw, 1981).

The networking of personal workstations in the 1980s made it attractive to develop groupware knowledge acquisition and modeling techniques that made overt individual and social constructions in groups working together on local area networks (Shaw and Gaines, 1991; Shaw and Gaines, 1993). The advent of the World Wide Web in the 1990s make it feasible to move these techniques to the web, and make widely available techniques for conceptual modeling that can be used in their own right or integrated with other applications.

This article describes experience in moving repertory grid elicitation and analysis onto the World Wide Web. WebGrid is an HTTP server variant of our RepGrid program (CPCS, 1993), the repertory grid subsystem of KSSn, an integrated suite of knowledge support system tools (Gaines and Shaw, 1993a; Gaines, 1994). WebGrid allows any web client supporting HTML level 2.0 to be used for repertory grid elicitation and analysis.

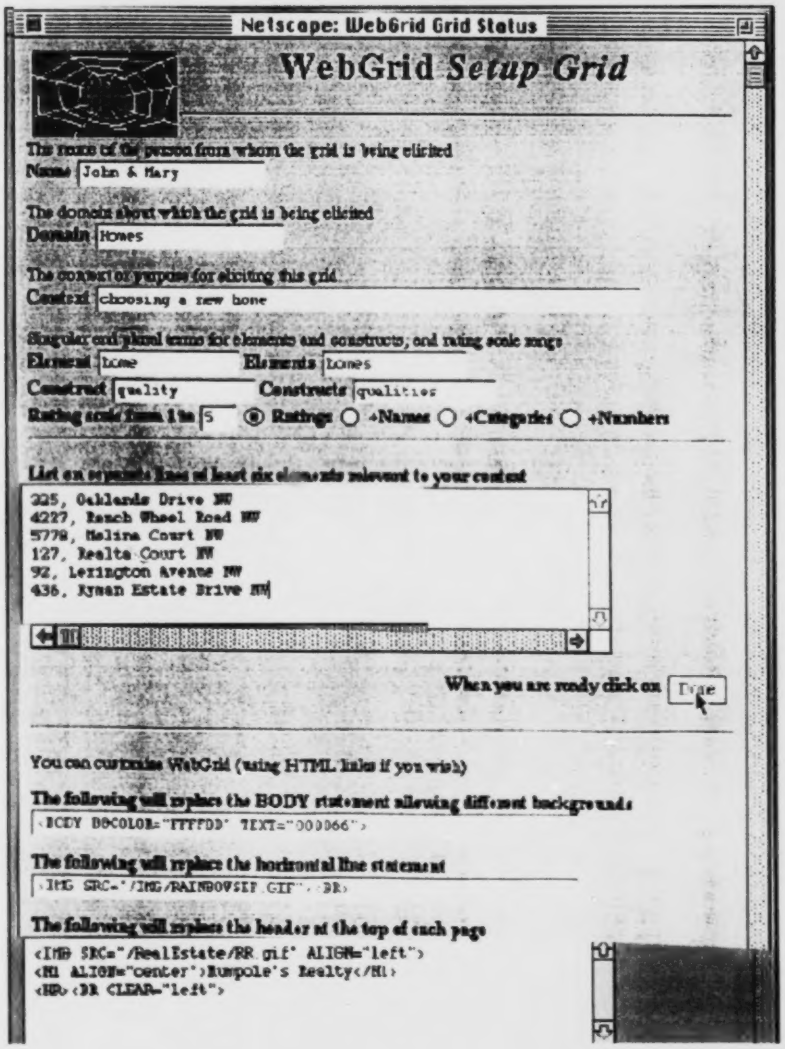
Most examples of repertory grid elicitation focus on specialist professional applications. However, the tools are generally applicable to any kind of decision making situation, and the application used to illustrate WebGrid in this paper is that of home purchase, a common experience typical of individual and group decision making.

2 WebGrid in Action

The simplest way to understand WebGrid is to use it on the web (http://tiger.cpsc.ucalgary.ca/WebGrid/), and the next simplest is to see someone else using it. This section shows WebGrid in action in a real estate decision making application.

2.1 WebGrid Set Up

The scenario is that John and Mary are looking to purchase a new home and the realtor is having them use WebGrid to build a conceptual model of the dimensions on which they are evaluating the homes. Figure 1 shows the initial set up screen which enables the realtor to customize WebGrid by specifying the general domain, the specific context, the terms for elements, the terms for constructs, an initial set of elements, and header and trailer HTML for each page which will replace the normal WebGrid defaults. This set up need only be done once for a particular application area.



WebGrid *Setup Grid*

The name of the person from whom the grid is being elicited
Name John & Mary

The domain about which the grid is being elicited
Domain Homes

The context or purpose for eliciting this grid
Context choosing a new home

Singular and plural terms for elements and constructs, and rating scale range
Element home **Elements** homes
Construct quality **Constructs** qualities
Rating scale from 1 to 5 ● Ratings ○ +Names ○ +Categories ○ +Numbers

List on separate lines at least six elements relevant to your context

```
325, Oaklands Drive NW
4227, Ranch Wheel Road NW
5778, Melina Court NW
127, Realta Court NW
92, Lexington Avenue NW
436, Ryman Estate Drive NW
```

When you are ready click on Done

You can customize WebGrid (using HTML links if you wish)

The following will replace the BODY statement allowing different backgrounds

```
<BODY BGCOLOR="FFFFDD" TEXT="000066">
```

The following will replace the horizontal line statement

```
<IMG SRC="/IMG/RAINBOWSIF.GIF"><BR>
```

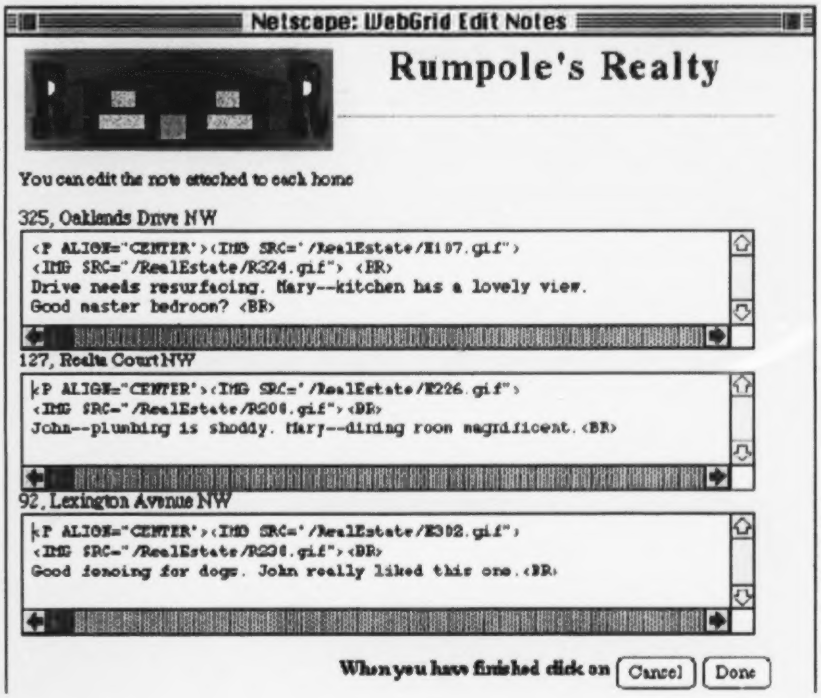
The following will replace the header at the top of each page

```
<IMG SRC="/RealEstate/RR.gif" ALIGN="left">
<H1 ALIGN="center">Rumpole's Realty</H1>
<HR><BR CLEAR="left">
```

BEST COPY AVAILABLE

Figure 1 Initial form for setting up a WebGrid elicitation

The HTML in the header text area second from bottom replaces the WebGrid icon and heading with one specific to the realtor as shown in Figure 2, so that John and Mary see pages customized with a corporate image. The initial set of elements that the realtor has entered consists of 6 homes that she has already shown to John and Mary. She has used the annotation form shown in Figure 2 to enter notes on each property using stored photographs of the homes and the personal comments that John and Mary have made when viewing the homes.



You can edit the note attached to each home

325, Oaklands Drive NW

```
<P ALIGN="CENTER'><IMG SRC='/RealEstate/H107.gif">
<IMG SRC="/RealEstate/R324.gif"> <BR>
Drive needs resurfacing. Mary--kitchen has a lovely view.
Good master bedroom? <BR>
```

127, Realta Court NW

```
<P ALIGN="CENTER'><IMG SRC='/RealEstate/H226.gif">
<IMG SRC-"/RealEstate/R208.gif"><BR>
John--plumbing is shoddy. Mary--dining room magnificent.<BR>
```

92, Lexington Avenue NW

```
<P ALIGN="CENTER'><IMG SRC='/RealEstate/H302.gif">
<IMG SRC-"/RealEstate/R236.gif"><BR>
Good fencing for dogs. John really liked this one.<BR>
```

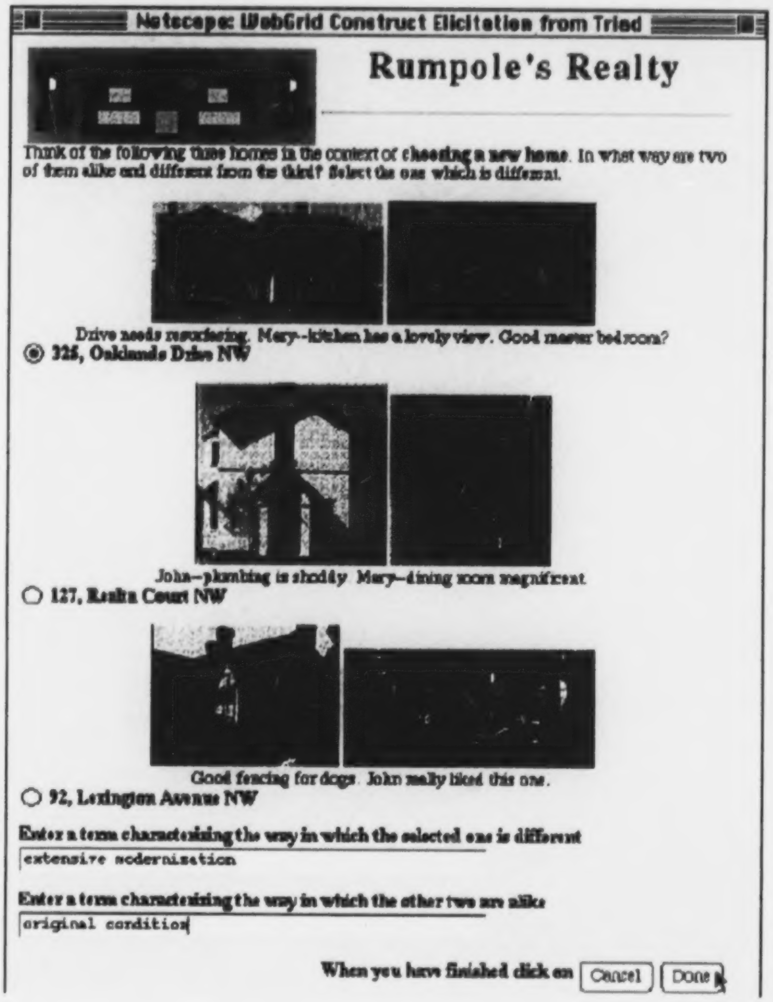
When you have finished click on

Figure 2 Annotating the elements with links to pictures and notes

2.2 WebGrid Triadic Elicitation Facilities

In practice, most of the customization of WebGrid, annotation of elements, and so on, in this type of application will be done through prestored forms and access to databases. The first screen that an end user sees is generally one of the form shown in Figure 3. This screen is generated by WebGrid using the standard repertory grid triadic elicitation technique of choosing 3 elements and asking in what way two are alike and different from the third. The user clicks in a radio button to select the one which is different, and types in his or her terms for the similarity and difference in the text boxes below. The similarity/difference define a distinction--they are the *poles* of a bipolar *construct* in Kelly's terminology.

BEST COPY AVAILABLE



Rumpole's Realty

Think of the following three homes in the context of **choosing a new home**. In what way are two of them alike and different from the third? Select the one which is different.

Drive needs resurfacing. Mary--kitchen has a lovely view. Good master bedroom?

● **325, Oaklands Drive NW**

John--plumbing is shoddy. Mary--dining room magnificent.

○ **127, Realta Court NW**

Good fencing for dogs. John really liked this one.

○ **92, Lexington Avenue NW**

Enter a term characterizing the way in which the selected one is different

extensive modernization

Enter a term characterizing the way in which the other two are alike

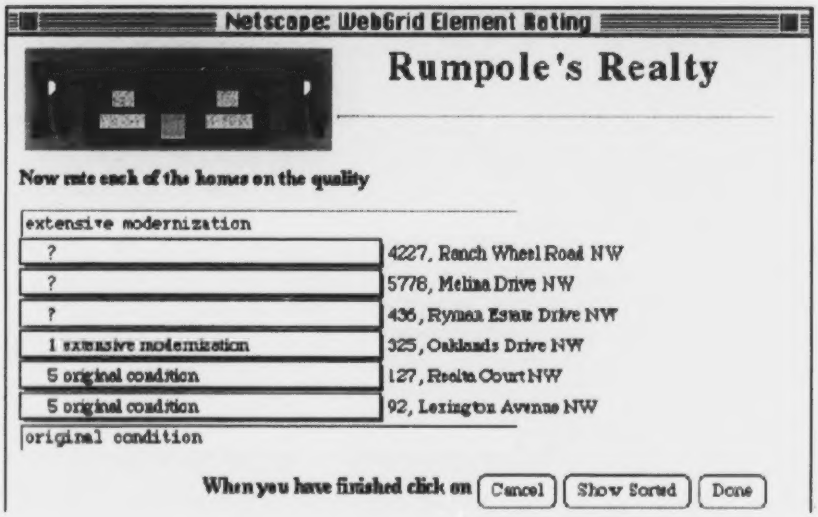
original condition

When you have finished click on

Figure 3 Triadic elicitation of a construct distinguishing elements

When the user clicks on "Done" WebGrid generates the form shown in Figure 4 which uses popup menus to allow all the elements to be distinguished in terms of the construct. It also allows the pole names to be changed or the original elements to be re-rated, if appropriate. WebGrid is designed to be non-modal, and allow users to easily edit previous entries at every stage.

BEST COPY AVAILABLE


Netscape: WebGrid Element Rating

Rumpole's Realty

Now rate each of the homes on the quality

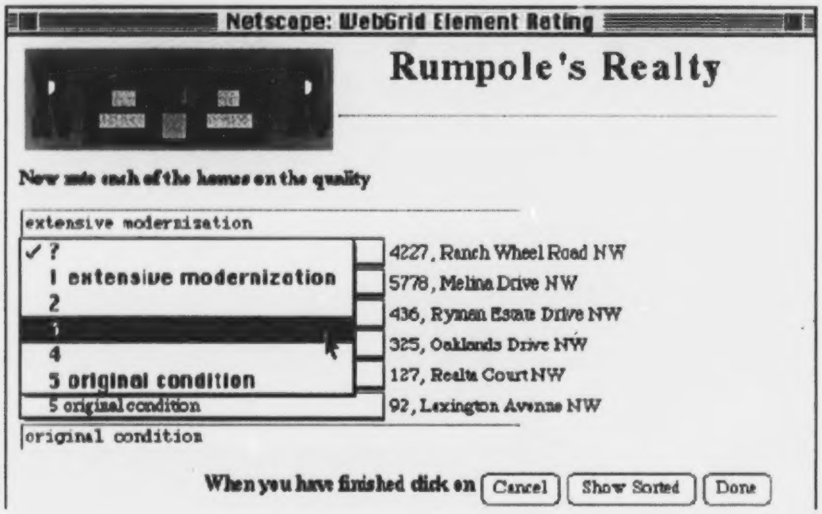
extensive modernization
? 4227, Ranch Wheel Road NW
? 5778, Melina Drive NW
? 436, Ryman Estate Drive NW
1 extensive modernization 325, Oaklands Drive NW
5 original condition 127, Realta Court NW
5 original condition 92, Lexington Avenue NW
original condition

When you have finished click on

Cancel
Show Sorted
Done

Figure 4 WebGrid form for rating elements on a construct

Figure 5 shows the popup menus being used as rating scales to position the elements along the construct.


Netscape: WebGrid Element Rating

Rumpole's Realty

Now rate each of the homes on the quality

extensive modernization

✓ ?
1 extensive modernization
2
3
4
5 original condition

4227, Ranch Wheel Road NW
5778, Melina Drive NW
436, Ryman Estate Drive NW
325, Oaklands Drive NW
127, Realta Court NW
5 original condition 92, Lexington Avenue NW
original condition

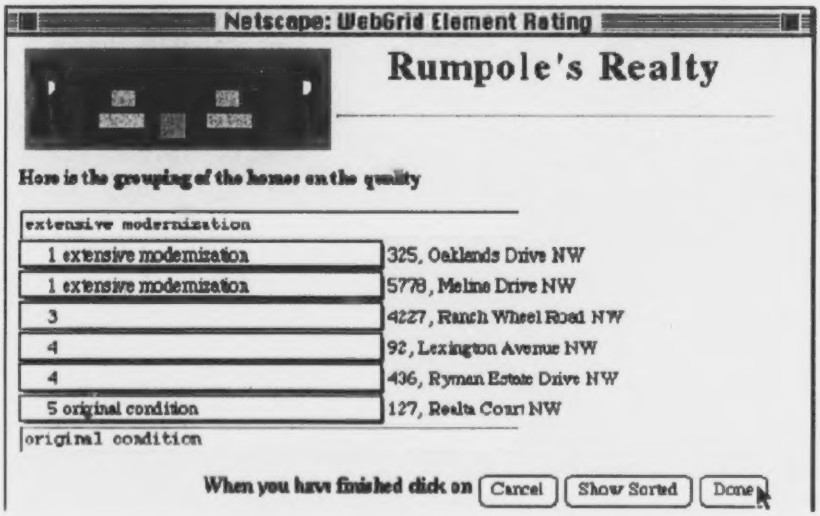
When you have finished click on

Cancel
Show Sorted
Done

Figure 5 Using popup menus to enter ratings

When all the elements have been rated the user can either click on "Done" to enter the data, or on "Show Sorted" to see the entered data positioned along a scale as shown in Figure 6. This is useful to check whether the order of the elements along the construct makes sense to the user.

BEST COPY AVAILABLE



Rumpole's Realty

Here is the grouping of the homes on the quality

extensive modernization

1 extensive modernization	325, Oaklands Drive NW
1 extensive modernization	5778, Meline Drive NW
3	4227, Ranch Wheel Road NW
4	92, Lexington Avenue NW
4	436, Ryman Estate Drive NW
5 original condition	127, Realta Court NW

original condition

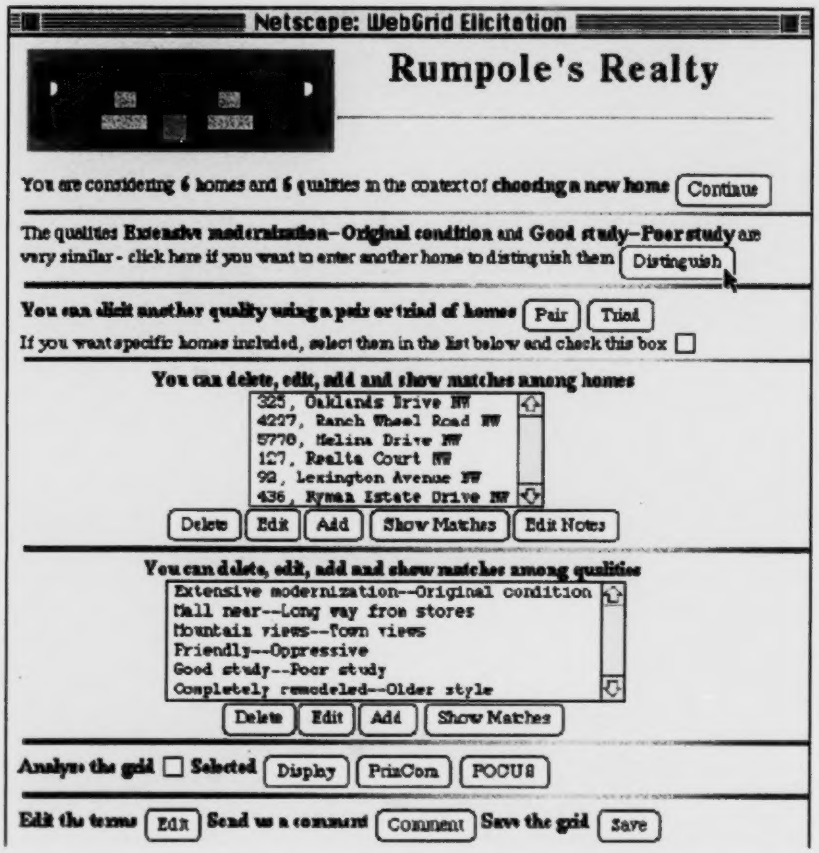
When you have finished click on

Figure 6 Showing the elements sorted along the construct

2.3 WebGrid Other Elicitation Facilities

When the user finally clicks on "Done" WebGrid generates its main status screen as shown in Figure 7 (after 6 constructs). This is divided into sections suggesting various actions that may be taken. WebGrid continually computes matches between elements and between constructs, and suggests that additional constructs or elements be entered to help reduce similarities which may be due to insufficient data having been entered. It allows the elements and constructs already entered to be edited, and it provides access to various analysis tools which are illustrated later.

BEST COPY AVAILABLE



Rumpole's Realty

You are considering **6** homes and **6** qualities in the context of **choosing a new home**

The qualities **Extensive modernization--Original condition** and **Good study--Poor study** are very similar - click here if you want to enter another home to distinguish them

You can elicit another quality using a pair or triad of homes

If you want specific homes included, select them in the list below and check this box ☐

You can delete, edit, add and show matches among homes

325, Oaklands Drive NW
4227, Ranch Wheel Road NW
5770, Melina Drive NW
127, Realta Court NW
92, Lexington Avenue NW
436, Ryman Estate Drive NW

You can delete, edit, add and show matches among qualities

Extensive modernization--Original condition
Mall near--Long way from stores
Mountain views--Town views
Friendly--Oppressive
Good study--Poor study
Completely remodeled--Older style

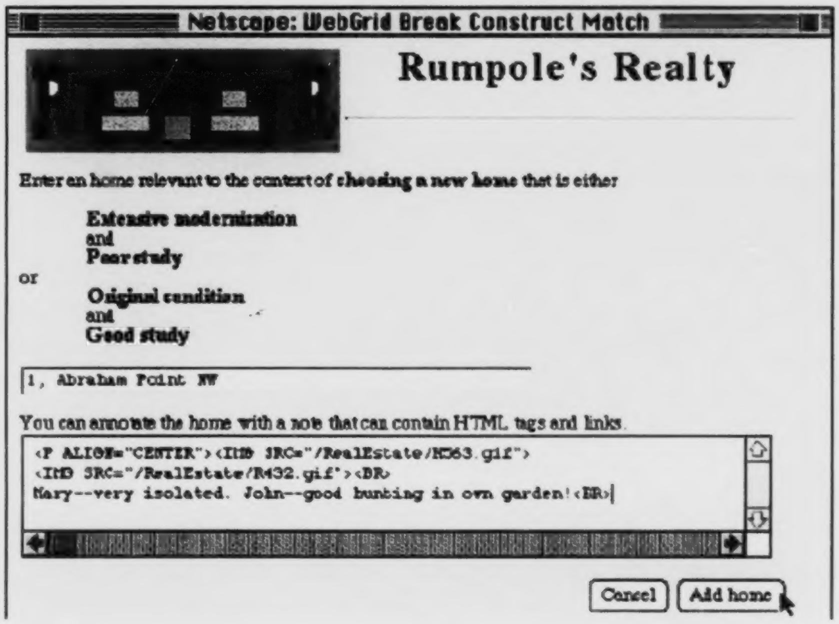
Analyse the grid ☐ **Selected**

Edit the terms **Send us a comment** **Save the grid**

Figure 7 Main status screen of WebGrid after 6 constructs have been entered

The suggestion at the top of the screen is that the constructs "Extensive modernization--Original condition" is closely matched with "Good study--Poor study". This may be a coincidence based on the small set of homes entered, or it may reflect a significant phenomenon such as modernization being improving study facilities. John and Mary decide to investigate this match, click on "Distinguish", and WebGrid generates the screen shown in Figure 8 which poses a very specific question--can they think of a home which is modernized and has a poor study or in original condition and has a good study.

BEST COPY AVAILABLE


Netscape: WebGrid Break Construct Match
Rumpole's Realty
Cancel
Add home

Enter an home relevant to the context of **choosing a new home** that is either

Extensive modernization
and
Poor study
or
Original condition
and
Good study

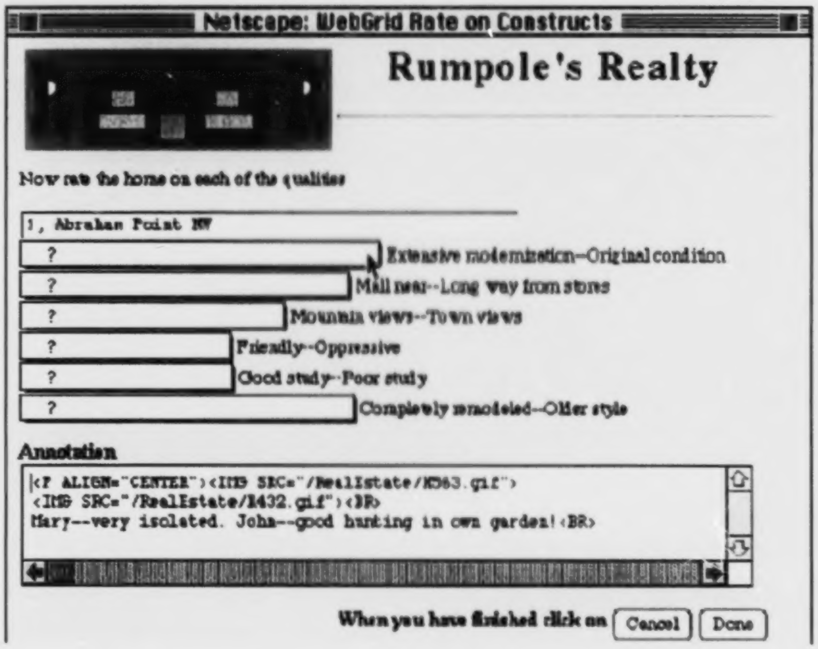
1, Abraham Point NW

You can annotate the home with a note that can contain HTML tags and links.

```
<P ALIGN="CENTER"><IMG SRC="/RealEstate/K363.gif">
<IMG SRC="/RealEstate/R432.gif"><BR>
Mary--very isolated. John--good hunting in own garden!<BR>
```

Figure 8 Eliciting a new element to break a construct match

This triggers the realtor into suggesting they view a particular home which she thinks may interest them that is modernized and has a study. They visit the home and add it as a new element as shown in Figure 8. WebGrid then generates the screen of Figure 9 asking them to rate the newly entered element on the existing constructs.


Netscape: WebGrid Rate on Constructs
Rumpole's Realty
Cancel
Done

Now rate the home on each of the qualities

1, Abraham Point NW

? Extensive modernization--Original condition
? Mall near--Long way from stores
? Mountain views--Town views
? Friendly--Oppressive
? Good study--Poor study
? Completely remodeled--Older style

Annotation

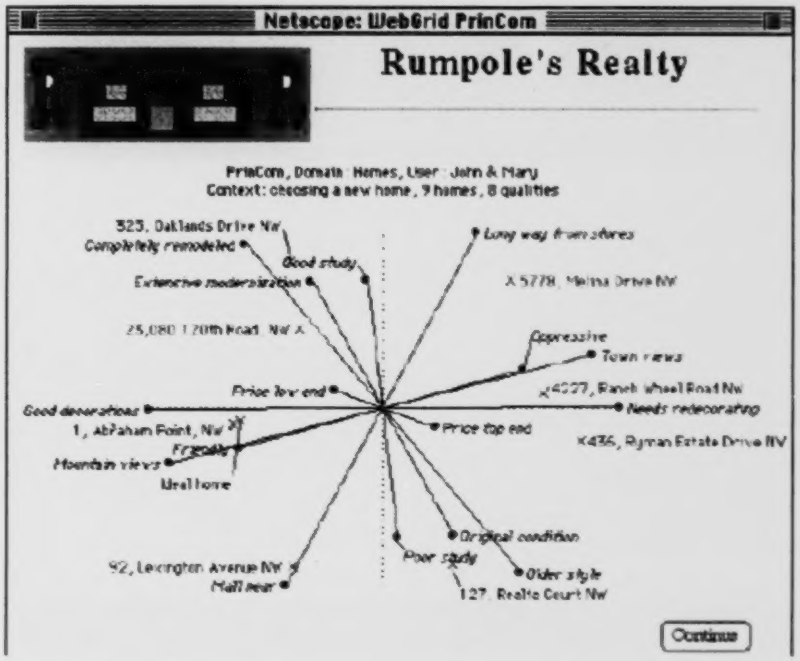
```
<P ALIGN="CENTER"><IMG SRC="/RealEstate/K363.gif">
<IMG SRC="/RealEstate/R432.gif"><BR>
Mary--very isolated. John--good hunting in own garden!<BR>
```

When you have finished click on

Figure 9 Rating a new element on existing constructs

2.4 WebGrid Modeling Facilities

As a final element, the realtor has John and Mary add the conceptual element, "Ideal home", and rate it on the existing constructs. This defines the preferred value, usually a "preferred pole", on each construct for WebGrid. The users now chose one of the analysis programs, *PrinCom*, to generate a map of the relations between constructs and elements. PrinCom does a principal components analysis of the grid data, attempting to represent it in as few dimensions as possible. WebGrid plots the first two components and returns them as a GIF as shown in Figure 10. It can be seen that there are two major clusters of dimensions characterized by "Friendly--Oppressive" and "Extensive modernization--Original condition". It can also be seen that "Ideal home" clusters most strongly with "1, Abraham Point".



Netscape: WebGrid PrinCom

Rumpole's Realty

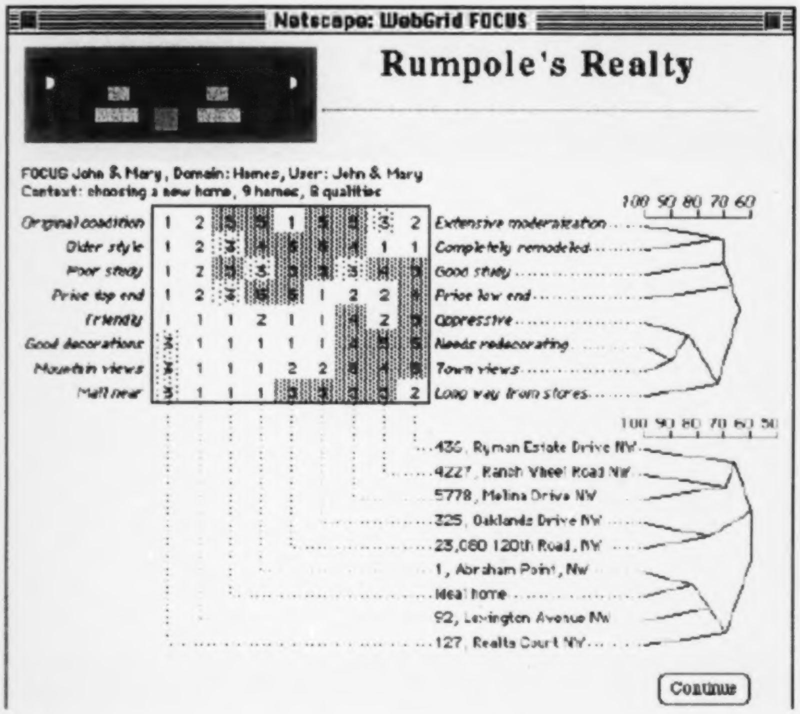
PrinCom, Domain: Homes, User: John & Mary
Context: choosing a new home, 9 homes, 8 qualities

323, Oaklands Drive NW
Completely remodeled
Good study
Extensive modernization
25,080 120th Road, NW
Long way from stores
5778, Melina Drive NW
Oppressive
Town views
4027, Ranch wheel Road NW
Price low end
Good decorations
Needs redecorating
1, Abraham Point, NW
Friendly
Mountain views
Ideal home
Price top end
436, Ryman Estate Drive NW
Original condition
Poor study
92, Lexington Avenue NW
Mall near
Older style
127, Realto Court NW
Continue

Figure 10 Map produced by PrinCom principal components analysis

The users also use the FOCUS analysis program to sort the grid so that similar elements and similar constructs are clustered together. WebGrid graphs the data and returns it as a GIF as shown in Figure 11. Again, it can be seen that "1, Abraham Point, NW" is closest to being the "Ideal home". Because the original grid data is shown, it is possible to see how this home and the others differ from the ideal.

BEST COPY AVAILABLE



FOCUS John & Mary, Domain: Homes, User: John & Mary
Context: choosing a new home, 9 homes, 8 qualities

Original condition	1	2	5	5	1	5	5	3	2	Extensive moderaization
Older style	1	2	3	4	5	5	4	1	1	Completely remodeled
Poor study	1	2	5	3	5	5	3	4	5	Good study
Price top end	1	2	3	5	5	1	2	2	4	Price low end
Friendly	1	1	1	2	1	1	4	2	5	Oppressive
Good decorations	3	1	1	1	1	1	4	5	5	Needs redecorating
Mountain views	3	1	1	1	2	2	4	4	5	Town views
Mall near	3	1	1	1	3	3	3	3	2	Long way from stores

436, Ryman Estate Drive NW
4227, Ranch Wheel Road NW
5778, Melina Drive NW
325, Oaklands Drive NW
23,080 120th Road, NW
1, Abraham Point, NW
Ideal home
92, Lexington Avenue NW
127, Realts Court NY

Figure 11 Clusters produced by FOCUS sorting analysis

The sequence of interaction in a real estate application shows how repertory grids may be used interactively across the web to elicit and model personal knowledge. This example is illustrative of a wide range of personal decision making applications, choosing among options, such as a new job or a new car.

3 Conclusions

When we commenced the development of WebGrid, knowing the limitations of HTML forms, we had no great expectations that the resultant tool would be attractive to use. We have spent many years carefully developing user interfaces with good human factors targeted on repertory grid elicitation, and knew the need for special-purpose graphical "widgets" that could not be emulated in HTML. We are enthusiastic users of World Wide Web for publication, and were interested to see what could be achieved with more interactive, collaborative tools, but felt that the technology might not yet be adequate for what we wanted to achieve. We were more than pleasantly surprised by the human factors and attractiveness of WebGrid. The free-form HTML documents with embedded widgets in many ways gave us more flexibility than we had experienced in the design of the original RepGrid interface, and interaction with WebGrid has proved natural to our user communities. From a programming perspective, HTML forms provide a cross-platform graphic user interface that is simple to prototype, easy to customize, and whose simple primitives help to support the basic human factors guidelines of *uniformity* and *consistency* (Gaines and Shaw, 1984).

One objective in the original WebGrid development was to preserve the stateless nature of the HTTP protocol. No data is stored at the server between transactions. In particular, this makes it possible for us to offer WebGrid as an open service on the web without being concerned about managing the storage of other people's data. However, we are also concerned to support collaborative communities as we have done with RepGridNet (Shaw and Gaines, 1991; Shaw and Gaines, 1993), where members can make their grids available to the community. This raises many problems of authorization, protection, and so on. It is not easy to design a system for the web that has open, easy access, yet avoids data loss or unwanted interference through careless or deliberate actions.

The real estate application was chosen to illustrate this article because it provides a good example of how a web-based system can be customized by end users. The web offers excellent facilities for dynamic hypermedia and for integrating these facilities with structured interaction to enhance general client-server programs.

WebGrid provides many additional facilities not illustrated in this paper. It allows a user to commence an elicitation based on another person's grid, either using just the elements in it and developing his or her own constructs, or using both elements and constructs but commencing with all the rating unknown (an "exchange" grid). In the first case the new grid may be compared with the original one to determine what *constructs* correspond in the two grids--the users may be making the same distinctions but giving them different names. In the second case the new grid may be compared with the original one to determine to what extent the *ratings* correspond in the two grids--the users may be making different distinctions but giving them the same names. These facilities are used with students to support collaborative learning (Shaw and Gaines, 1995) and with groups of experts to explore difference in the use of constructs and terminology (Shaw and Gaines, 1989).

WebGrid also provides the expert system development tools based on repertory grids that have been used extensively in knowledge-based system development (Gaines and Shaw, 1993a; Gaines and Shaw, 1993b). One option in Figure 1 is to allow a wider range of data types including categorical data, numbers, dates, and so on, and when this is selected additional choices appear on WebGrid elicitation and analysis screens.

We are using WebGrid routinely to support graduate and undergraduate teaching (Shaw and Gaines, 1996). It is also available on the Internet, and in the period July-December 1995 WebGrid was accessed from 674 different sites in 30 countries. The web is essentially an anonymous medium and we do not track the activities of outside users. However, occasionally users contact us to discuss their applications and, if the local network goes down, we receive mail from serious users reporting the problem and requesting notification of when the knowledge acquisition programs will be available again. An interesting example of serendipitous use was a masters student in Holland studying operators' models of nuclear reactors who came across the grid elicitation system on the web in September, used it with his experts to elicit their conceptual models, and included these in an additional chapter in his thesis for examination in October--the web certainly accelerates research processes.

Acknowledgments

This work was funded in part by the Natural Sciences and Engineering Research Council of Canada.

References

CPCS (1993). **RepGrid 2 Manual**. Centre for Person Computer Studies, 3019 Underhill Drive NW, Calgary, AB, Canada T2N 4E4.

Gaines, B.R. (1994). Class library implementation of an open architecture knowledge support system. **International Journal Human-Computer Studies 41**(1/2) 59-107.

Gaines, B.R. and Shaw, M.L.G. (1984). **The Art of Computer Conversation**. New Jersey, Prentice-Hall.

Gaines, B.R. and Shaw, M.L.G. (1993a). Basing knowledge acquisition tools in personal construct psychology. **Knowledge Engineering Review 8**(1) 49-85.

Gaines, B.R. and Shaw, M.L.G. (1993b). Eliciting knowledge and transferring it effectively to a knowledge-based systems. **IEEE Transactions on Knowledge and Data Engineering 5**(1) 4-14.

Kelly, G.A. (1955). **The Psychology of Personal Constructs**. New York, Norton.

Shaw, M.L.G. (1980). **On Becoming A Personal Scientist: Interactive Computer Elicitation of Personal Models Of The World**. London, Academic Press.

Shaw, M.L.G., Ed. (1981). **Recent Advances in Personal Construct Technology**. London, Academic Press.

Shaw, M.L.G. and Gaines, B.R. (1989). A methodology for recognizing conflict, correspondence, consensus and contrast in a knowledge acquisition system. **Knowledge Acquisition 1**(4) 341-363.

Shaw, M.L.G. and Gaines, B.R. (1991). Supporting personal networking through computer networking. **Proceedings of CHI'91: Human Factors in Computing Systems**. pp.437-438. New York, ACM.

Shaw, M.L.G. and Gaines, B.R. (1993). Group knowledge elicitation over networks. Bramer, M.A. and Macintosh, A.L., Ed. **Research and Development in Expert Systems X. Proceedings of British Computer Society Expert Systems Conference**. pp.43-62. Cambridge, UK, Cambridge University Press.

Shaw, M.L.G. and Gaines, B.R. (1995). Comparing constructions through the web. Schnase, J.L. and Cunnius, E.L., Ed. **Proceedings of CSCL95: Computer Support for Collaborative Learning**. pp.300-307. Mahwah, New Jersey, Lawrence Erlbaum.

Shaw, M.L.G. and Gaines, B.R. (1996). Experience with the learning web. Carlson, P. & Makedon, F., Ed. **Proceedings of ED-TELECOM 96 : World Conference on Educational Telecommunications**. pp.320-325. Charlottesville, VA, Association for the Advancement of Computing in Education.

KSI Page

gaines@cpsc.ucalgary.ca 16-Jul-96

Media, Multimedia and Meaningful Language Learning: A Review of the Literature

Peter Shea
Department of Educational Theory and Practice
University at Albany, State University of New York
USA
PS7669@CNSVAX.ALBANY.EDU

Abstract: This review examines issues surrounding the effects of media on language learning. Ausubel's notions of meaningful and non-meaningful learning are applied to the theories and practices surrounding language acquisition and media use in an attempt to discern whether the media employed in language teaching may have an effect on the cognitive processes involved in language learning. Various conceptualizations of language acquisitions are analyzed to determine if they influence media application and cognitive outcomes. Studies are presented which support the position that media, through their provision of different and potentially complementary symbol systems, can have positive effects on learning outcomes. It is concluded that the linguistic and pedagogic theories underlying media use tend to enable or constrain that effect.

The difference between meaningful and non-meaningful learning is an essential one and has been a question of interest in the field of language learning. The question of the effects of media on learning has also been hotly contested (Clark 1983, Kozma, 1991). Several commentators have noted that traditional uses of media in language learning have failed to result in meaningful learning (Rivers, 1964, Richards & Rodgers, 1991). One aspect of language learning research that has not been investigated fully involves an inquiry into the conditions under which media can contribute to meaningful language learning. This review attempts to provide some tentative answers to these questions through an analysis and integration of theoretical and empirical data.

Rote and Meaningful Learning

David Ausubel (1966) makes a distinction between rote and meaningful learning that is useful to a inquiry regarding the value of media for language learning. The author describes rote learning as the process of acquiring material as "discrete and relatively isolated entities that are relatable to cognitive structure only in an arbitrary or verbatim fashion, not permitting the establishment of relationships (Ausubel, 1968)." In this type of learning items are stored in ways that have little or no association with existing cognitive structure. For example, in language learning students are often able to produce canned or formulaic expressions without meaningfully understanding. While such expressions can contribute to language learning (Chesterfield 1985, Oxford 1990), true fluency requires meaningful understanding.

Ausubel contrasts rote learning with meaningful learning which is described as a process of relating and anchoring new material to relevant, established entities in cognitive structure. To achieve this requires that two criteria be met. First, the student must have a meaningful learning set i.e. a disposition to relate material in a nonarbitrary way to existing cognitive structure. Second, the task which the student performs in the learning process must itself be potentially meaningful to the student, i.e. relatable in a non-verbatim fashion (Ausubel, 1966). In learning, material is meaningful if it is subsumable i.e., if the learner has some relevant cognitive structures to which the material can be connected and a disposition to do so. Through a process of systematically relating material with existing structure and making those structures denser and more inclusive material is retained. The significance of the distinction between rote and meaningful learning lies in retention. Items that are learned by rote are far more transient than those learned meaningfully.

Theories of Language and Meaningful Learning

While the theory of language beneath Communicative Language Teaching (CLT) is fairly explicit and grounded in the notion of communicative competence, the theory of learning in this approach is not so apparent. Richards and Rodgers (1991) assert that three principals appear to be implicit in the suggested learning activities of CLT :

One such element might be described as the communication principle. Activities that involve real communication promote learning. A second element is the task principle: Activities in which language is used for carrying out meaningful tasks promote learning. A third element is the meaningfulness principle: Language that is meaningful to the learner support the learning process. Learning activities are consequently selected according to how well they engage the leaner in meaningful and authentic language use, rather than mechanical practice of language patterns. (p72)

Whereas earlier approaches, based on behaviorism, focused primarily on grammatical and lexical capacity, CLT endeavors to attend to all of the competencies through focusing on the communicative function of language. Ausubel's notions of meaningful and non-meaningful learning and subsumption are supportive of such an approach.

Learning with media

Kozma (1991) raises the question as to whether media influence learning. He presents a framework in which media are conceived as potentially complimentary to the processes involved in meaningful learning, i.e. the creation of knowledge through an integration of new content with existing cognitive structures. He contends that the symbol sytem characteristic of a media may facilitate learning if the instructional method exploits the media's capabilities to provide representations which are effective and relevant to the task and situation (p. 5). This framework stands in opposition to a view of media as "vehicles of delivery" which have no direct effect on the learning process (Clark, 1983 p. 445).

What do studies of media and language learning suggest regarding the role of media in meaningful language learning i.e. the acquisition of commucicative competence? Are media merely a vehicle for delivering instruction or are there attributes to certain media which more effectively facilitate the subsumptive cognitive processes involved in acquiring communicative competence? Finally can media be used to meet the goals of creating learning activities which engage the leaner in meaningful and authentic language use, rather than mechanical practice of language patterns?

Non-meaningful language learning and media

A vehicle for delivering instruction: the audiolingual method The Audiolingual approach, probably more than any language teaching method, has utilized media as a key component. The language laboratory is an invention of the proponents of this method. Tens of thousands of students have spent tens of thousands of hours listening to the audio tapes that have been a staple of foreign language instructional materials. This method of instruction has however, met with remarkably limited success (Rivers, 1964). An easy conclusion to draw is that media may not have a very powerful impact on the process of language learning. It even seems possible that certain features of media may interact in negative ways with the characteristics of second language learners and actually impede the cognitive processes involved in language learning.

Meaningful language learning and media

While studies of second language learners uses of media are rare, a number of studies have addressed questions regarding how the characteristics of audio and audiovisual media might affect learning in the student's native language. These studies often compare the effects of different media on the learning of children and adults. It might be argued that the more limited cognitive resources which children must cope with are in some ways analogous to the strained cognitive processes which are typical of the second language learning experience. Specifically, the limits of attention and smaller knowledge bases of children may be considered comparable to the heavier cognitive load and limited knowledge bases which confront adult second language learners. Because results of first language studies are potentially analogous and due to the scarcity of data involving second language learners, studies which investigated the effects of media on native speakers language learning processes were selected for inclusion in this review.

One question which arises in a discussion of media and language learning involves learner characteristics. The cognitive demands of language learning can be great. Attending to and attempting to decode a foreign language requires that mental resources be employed in ways that can easily result in misunderstanding. New students are especially unlikely to have the relevant cognitive structures necessary to make meaning. In order for meaningful learning to occur a basic prerequisite is that the medium facilitate rather than inhibit the cognitive processes involved in the learning. Is it possible that the added cognitive strain of attending to certain media, for example those characterized by their engagement of multiple channels of perception such as networked multimedia environments, may result in cognitive overload and comprehension failure? The review of relevant studies which follows attempts to address this question.

Dual Symbol Systems

A number of studies (Beagles-Roos, J. & Gat, I. 1983, Gibbons, J., Anderson, D., Smith, R., Field, D.,& Fischer, C. 1986, Pezdek, K. & Hartman, E. 1983, Pezdek, K. & Stevens, E. 1984, Hayes, D., Kelly, S., & Mandel, M. 1986) investigated the question of whether the simultaneous presentation of audio and visual information taxes cognition at the expense of comprehension. One hypothesis in this debate is that, because ny symbol system can carry meaning, restricting the flow of information to a single channel reduces the amount of data which must be processed, allowing the learner to more effectively focus on the significance of the message. A report by Gibbons (1986) is somewhat representative of studies in this area. Ninety-six children aged 4.5 and 7.5 were presented stories in either audio or audiovisual form to investigate the hypothesis that visually presented content is more memorable for children. The children were asked to recall or reconstruct the stories. Audiovisual input led to better memory for explicitly presented action as well as provoking more elaborations. The younger children were more likely to remember actions than utterences regardless of medium.

Similar to Gibbons, none of the other investigators found that the simultaneous presence of two symbol systems resulted in reduced comprehension for the younger and hence more cognitively challenged subjects. In most of the studies (Baggett and Ehrenfeucht 1982, Beagles-Roos, J. & Gat, I. 1983, Pezdek, K. & Stevens, E.

1984) the authors concluded that simultaneously presented symbol systems resulted in more recall than presentations using text only, audio only or video only. In these studies oral reconstructions and ordering tasks were used to measure recall and hence comprehension. These tests can be viewed as measures of discourse competence since the students had to combine grammatical forms and meanings to achieve a unified or written text. Grammatical competence is also demonstrated in these tasks.

What is the relationship between the symbol system which a medium employs and the process of subsumption involved in the acquisition of communicative competence in a foreign language? Remembering Ausubel's definition of meaning is useful in answering the question: "Meaning is a clearly articulated and precisely conscious experience that emerges when potentially meaningful signs, symbols, [italics added] concepts, or propositions are related to and incorporated within a given individual's cognitive structure on a nonarbitrary and substantive basis." (cited in Brown, p. 79) The results of the studies reported above may be accounted for through an analysis of the nature of the symbol systems employed by the media. Children were able to remember more and produce more accurate retellings when information was presented audiovisually because such presentations provide more opportunities for incorporating material into relevant cognitive structure. Bagget (1989) suggests that visual representations are "bushier", i.e. they carry more information. This common sense notion is captured in the maxim - a picture is worth a thousand words.

Ausubel (1966) would also predict that children in the elementary school years would benefit more from visual i.e. concrete expressions of concepts than adults who are more adept at handling oral i.e. abstract representations (p.166). In general, audiovisual presentations, supplying additive and complementary information of two symbol systems, are probably more beneficial to students with less domain knowledge , while audio may be sufficient for those with more content knowledge (Kozma 1991, p. 14). Furthermore, information essential to the instruction of sociolinguistic competence, i.e. data that informs hypotheses regarding the relationship and attendant levels of politeness, register and formality of conversants, may be more difficult to interpret in purely aural presentations of language.

Three Symbol Systems

A number of other studies investigated the question of whether the simultaneous presentation of three symbol systems would impede or facilitate language learning. The subjects in most of these studies were second language learners. The media under investigation in this research was captioned or subtitled video. Again a primary question was whether the addition of another channel of information requiring the decoding of yet another symbol system would result in cognitive overload for students who are already engaged in a mentally taxing process.

Studies which looked at the effects of captioning on language learning (Borras, I. 1994, Garza, T., 1991, Koskinen, P. & Newman, S., 1991, Bean, R. M. & Wilson, R. M., 1989, Markham, P., 1989, Vanderplank, R., 1988,) used designs which compared groups who were tested on materials which were presented using either print, video, or video with captions. The design of the study reported by Garza (1991) is in many ways typical of this literature. In this analysis fourty students of Russian as a Foreign Language and seventy students of English as a Second Language participated in an experimental study to assess the effects of captioned video on language learning. Half the students of Russian watched videos with captions and half watched videos without captions. The students of English underwent the same treatments. On tests of language comprehension the groups that saw captioned video demonstrated a mean gain of 75.2 percent of correct answers, a mean decrease of 61.16 percent of incorrect answers and a mean decrease of 83.76 percent in unanswered questions relative to groups who saw the same video without captions.

While the gain scores reported by Garza are atypically high, in none of the studies did the presence of captions result in lower scores on tests of comprehension. In the other studies in which the subjects were second language learners, (Borras, I., 1994, Koskinen, P. & Newman, S., 1991, Markham, P., 1989) the presence of a captions improved comprehension relative to the other treatments. It would appear from these reports that the addition of a more stable sysmbol system, text, to a transient medium, video, results in increased meaningfulness of the language. Garza (1991) concluded that the captions provide a bridge between the student's strong reading ability and weaker listening skills (p. 259).

It should probably not be surprising that even second language learners are able to attend to multiple symbol systems simultaneously when we consider the results of Pezdek and Hartman (1984). In this study sixty five-year-old children watched a videotape in one of three conditions, A) with toys available to play with, B) with a record playing, C) with no such distractors. The dependent measures were percent of time that subject attended to the televison and recall accuracy on comprehension questions of the material viewed. Although attention to the televison varied by treatment, recall comprehension was not significantly different across treatment groups (p. 216) . Even five-year-old children were able to effectively distribute their attention such that they could process auditory and visual information while performing other tasks. The authors conclude that children learn a sophisticated cognitive processing strategy for television viewing at a very young age. It appears that second language learners are equally prepared for the task of processing multiple channels of information.

In two of the studies of captioning (Bean, 1989, Reese, 1983) the presence of captioning did not result in favorable effects on the learning. In both of these studies, subjects were using their native language. In the Bean article, subjects were adults learning to read. The author used captioning to determine whether subjects could spontaneously learn new vocabulary. Subjects were assigned to one of three conditions: 1) Closed captioned with instruction (n=9), 2) Closed caption without instruction (n=7), 3) Script with instruction (n=8). Instruction entailed discussing essential vocabulary with the students before presenting the material. Two weaknesses seem evident in this study: 1)The sample size is too small and 2)the expectation that students who cannot read will learn vocabulary spontaneously is too large. It appears that for those who have not adequately learned the skill of decoding letters to form words captioned video is not immediately beneficial.

In Reese (1983) subjects were assigned to one of four conditions in which news stories were presented either in their original form, or with the addition of subtitles. In two conditions the audio and video were mismatched. The resulting groups were. 1)Audio and video match, no captions, 2)Audio and video mismatch, no captions, 3)Audio and video match, captions present, 4)Audio and video mismatch, captions present. Subjects were told to evaluate the programs on a number of criteria including interest, clarity, accuracy, etc. The subjects were not told that they would be tested or that they should try to recall as much information as possible. In analyzing data the author noted that subjects reacted negatively to the "captions for the hearing impaired (p14)." These results seem to suggest that captions were viewed not as a potential aid to recall, (the subjects were unaware of their purpose in viewing), but as a compensatory meant for others. Kozma suggests that, "the perceptions students have about a medium and the purpose they have for viewing influence the amount of effort they put into the processing of the message and consequently, the depth of their understanding of the story (p. 13)." In Reese's study it appears that the students were misinformed about their purpose and had negative perceptions about the medium. In studies in which students were aware of their purpose and felt positive about the medium (Borras, I. 1994, Garza, T., 1991, Koskinen, P. & Newman, S., 1991) the presence of subtitles had a positive effect.

It seems clear that the theory of learning underlying the use of media will have an effect as to whether the characteristics of the media are effectively exploited. Early teaching methods which conceptualized learning as behavior tended to view media as a means to deliver stimuli rather than to engage cognitive processes. The failure of the Audiolingual approach is probably mainly due to the ways the media were used, the most common of which were to present repetition drills for the purpose of rote memorization. Using media in this fashion is not likely to contribute to the subsumptive processes referred to by Ausubel. Studies mentioned here (Beagles-Roos, J. & Gat, I. 1983, Gibbons, J., Anderson, D., Smith, R., Field, D.,& Fischer, C. 1986, Pezdek, K. & Hartman, E. 1983, Pezdek, K. & Stevens, E. 1984, Hayes, D., Kelly, S., & Mandel, M. 1986) also suggest that the medium audiolingualism often employed, audio tape, was among the least salient in engaging the cognitive processes involved in meaningful language learning.

Studies review here also tend to support Kozma's contention that the symbol systems which characterize media can have an effect on learning. Although the studies were more concerned with the media than the method of instruction, the results tend to suggest that media can play a role in the teaching of the various abilities which compose communicative competence. In accordance with the principles underlying CLT media should be used to enhance meaningfulness and communication which in turn facilitate the growth of language competencies. One of the most useful qualities of media for second language learning seems to be their potential in supporting the meaningfulness principle outlined above. The mode of presentation of certain media appear to make the language more meaningful to the learner, thus supporting the learning process.

As technological advances in computing increasingly integrate the various symbol systems utilized in communication, knowledge of the effects of the media which employ these symbol systems becomes more important to educators. Networked environments will soon provide access to high quality recorded and live digital video and text. These advances are blurring the lines between multimedia and telecommunications. While studies which investigated the effects of audio, audiovisual and captioned audiovisual presentaions support the idea that media can have differential effects on learning, researchers are only just beginning to investigate the possible effects of combining networked computing environments and access to multi-modal presentaions on learning. Some potential advantages to networked learning environments for language learning are their access to authentic environments, capacity for engaging motivation, and support for collaborative activities. Of these, access to authentic environments and the enhanced potential for interaction they allow may prove most useful if we consider the principals which underly second language acquisition.

Note: This is a much edited version of a 30 page review. Interested parties may recieve a copy of the full paper by sending e-mail to the author at: PS7669@CNSVAX.ALBANY.EDU

Mediation In The Internet

Jenny Shearer
HyperMedia Unit & Political Studies Department, University of Auckland, New Zealand.
jshearer@cs.auckland.ac.nz

Abstract: This paper offers a framework for mediation amoung oppositional policy positions based on a concept of the Internet as a self-regulating environment. It employs the structure of organisations which have a mandate to serve the interests of the Internet community. Organisations such as the Web Society (http://info.websoc.org) could mediate amoung oppositional policy positions which threaten the future of the Internet. National and commercial interests are developing policies which are seen as damaging by the Internet community, and equally, national, cultural, and commercial interests see some activities carried out on the Net as damaging to them. It is postulated that an international Internet mediation committee could make findings based on assessments of problems referred to it, using a set of principles based on freedom of use and development of the Web. The paper presents cases of interventions already taking place on an ad hoc basis within the Internet, and attempts to formalise the process.

Introduction

The current role of existing organisations (eg WebSoc, which sees itself as an "automobile association" of the Web) may be readily extended to provide a "mediation" service which could work towards finding solutions to a number of major problems which pose a significant threat to the future development of the Internet. This initiative could enhance progress towards creating a stable and free environment for development of; a politically unfettered electronic news media, comprehensive content bases, and rapid technical progress in the Internet. The " mediation" concept would assume a common identification by Web users with their supra-national community, and solutions could be established using a set of principles established on a global, rather than a national basis. The ethical problems raised by such an attempt may be, to a point, bypassed by an essentially pragmatic approach already adopted by organisations like WebSoc - that the important issue to be addressed is to make the Web workable for its community. Several important areas of conflict have been chosen as examples for the purpose of this paper, which is an attempt to set up a framework for Internet mediation. They are cryptography regulation, an area where Governments may move to gain access to private communications, pornography, a social issue in which consensus cannot be gained, either within countries or between countries, and the legal/technical problems raised by software patents and copyright litigation.

In order for it to progress, principles of freedom of use and development of the Internet must be developed. Step two is to justify when these principles may be subject to intervention - for example, when the harm done by complete freedom in some instance outweighs the importance of upholding the principle. Step three is to establish how the interventions may be carried out, and by whom. This paper "puts up" a set of principles and the perameters of interventions which may be considered acceptable to the Internet community, on the basis that these interventions may be viewed as self-regulation which will protect the community against solutions imposed by powerful national interests. An international committee could mediate individual cases or problems, and interventions carried out by the Internet community.

Interventions by community members acting either individually, or in groups, already occurr on an ad hoc basis. A valuable lesson to be learned from these "self-regulating" activities within the Net is that they frequently involve exposure of the problem, identification of the person creating the problem, and education of the wider community about the issues. Solutions and active interventions developing from a mediation committee's findings would be largely Net specific: that is, would employ techniques such as exposure of issues, identification of users involved in covert "illegal" activities, education of the Internet community about the issues, and social pressure on perceived "offenders". This focuses on the key issue of accountability, which so far has not been addressed by the Internet community. However, solutions could include international lobbying and co-operation with international policing agencies. The mechanism would be voluntary - if there was significant dissent with the Committee's findings amoung its community, the findings would not be actioned.

Schwerin advocates mediation as ìempoweringî for communities and individuals taking part. This is a key concept for the fragile ìcommunityî of the Internet, which is at risk of being overwhelmed by nation-based legislation while still in a developmental stage. Mediation is a key concept of transformational politics, which holds that major changes are needed to the existing economic, social, and political systems in order to bring about greater social justice, human growth, and effectiveness.

ì...these transformations should be evolutionary, not revolutionary, and they should ideally be nonviolent and noncoercive.î [Schwerin, 1995]

Schwerin warns that the ideology and practice of mediation is largely untested by academic standards, and that central concepts such as ìempowermentî have not been well-defined. However, balanced against that, mediation is becoming widely practised in America as an alternative to the legal system, and is perceived as having advantages of ìlower cost, greater accessibility to dispute resolution services, and the possibility of creating mutually satisfactory, durable agreements that are voluntarily crafted by the disputants.î [Schwerin,1995]

Mediation within the Internet community may be perceived as working alongside existing legal systems, both national and international. It has key advantages of being established on global principles rather than national, the process is committed to information sharing and education, and the process would have flexibility in dealing with problems as they impact within the Internet community, or with problems emanating from without (for example cryptography regulation). Mediators, selected from within the ranks of respected Internet ìgurusî, would have an understanding of the causes of the problems presented. Internet ìgurusî already hold court within the newsgroups and other areas of the Internet, and use social pressure against people who offend against existing mores. A structured mediation process would empower the best of these mediators, by giving them a forum and official community support, and a baseline structure of principle known to be widely accepted by the Internet community.

1 Establishing general principles: examples:

Freedom of Speech.

This may be cited as a "traditional" Internet value.

Access.

Universal access to the Internet and to its databases. (The networked technology of the Internet implies that universality will cause the network to work with maximum efficiency as a communications medium.)

Privacy. Privacy of communications. This may be cited as a traditional Internet value.

Freedom of Development. Freedom to develop software, content bases etc without undue barriers of cost or legal constraint, has been a founding principle of the Internet, which has been developed through the sharing of expertise.

It would be tempting to add opposition to abuse of human rights.

The building of a ìBill of Electronic Rights and Ethicsî, which is intended to ìqualifyî legislation, rather than to become legislation itself, is supported by the Electronic Freedom Frontier and other Internet lobby groups. A document [Electronic Rights and Ethics, 1996] puts forward ìrightsî of Internet users, under headings of communication, privacy, jurisdiction, access, and administration, and also proposes an ethical position to be supported by the Internet community. These ethics are expressed in terms of ìtoleration î, trust, consideration, and regulation, and apply directly to Internet activity, for example ìit is unethical to misrepresent yourself, your observations and opinions, or the expressions of othersî and ìit is unethical to be inconsiderate of the costs of network bandwidth and storage space.î These expressions of ethics and rights illustrate the development of an Internet ideology, and are a valuable attempt to articulate policy positions held by many Internet users. This value may be expressed through a mediation process, where such social mores or perceived rights provide a valuable guide for discussion and justification of policy positions.

2 Freedom of Speech and the Justification for Interventions

A key value of the Internet is freedom of speech. This was recently reinforced by the actions of a Canadian researcher who organised the setting up of miror-sites ("Zundelsites") to enabled Germans to read censored neo-Nazi material. Deutsche Telekom blocked acess to a small Santa Cruz company that maintains Zundel's World Wide Web site. Rich Graves set up a mirror site for Zundel's material, even though he disagrees with Zundel's views, and has put up arguements against them. Other mirror sites were established, and the attempt at censorship was dropped. The San Jose Mercury News reported:

"Declan McCullagh, a recent graduate from Carnegie Mellon who created the mirror site there, said censoring Zundel's message just lends credence to his arguement that there is a conspiracy preventing "the myths" about the holocaust from being revealed.

"The best way to respond to lies and deception is by rebuttal and refutation," he said. "The best efforts of a major Western democracy (at censorship) were thwarted by a couple of college students with an hour or two of spare time on the weekend." [San Jose Mercury News, 1996]

This action provides a blueprint for the successful handling of "offensive" material within the Internet. It involved exposure of the material to public view, criticism of the material, and caused the writer to be accountable to the public. People who set up mirror sites supported the principle of freedom of speech, but through the intervention of adding critical commentary and forcing the author of it (Ernst Zundel) to be exposed to public questioning, also gave the public an opportunity to assess the truth or otherwise of the allegations made by Zundel.

3 Cryptography Regulation

The interest of Governments in regulating for key forfeiture systems and surveillance holds insufficient weight against the rights of citizens to privacy of communications. [Shearer, Gutmann, 1996]

The notion of a right to privacy of citizens in their communications should be seen in the context of an international movement by governments towards regulation of cryptography, and consideration of key forfeiture systems in national cryptography use. The right to privacy in computer-based communications is an issue of major importance, assuring freedom of the individual in national and global communications. Regulation and control of cryptography use in the Internet by national governments may lead to an imbalance in the citizen/government power relationship, with consequences including unprecedented surveillance of citizens, disruption of international commerce due to lack of powerful cryptography (and lack of standardisation), human rights abuses by less democratic or non-democratic governments, and limiting of the political potential of an Internet global system. It is conceivable that a mediation committee would find that the Internet community strenuously opposes cryptography regulation. In the United States, a National Research Council Committee studying the cryptography issue has recommended that cryptography controls be progressively relaxed, and that the Government should stop pushing escrowed encryption. The Governmentís efforts in this direction are in the third edition , legislation known as ìClipper 111.î In view of political activity in the United States and in other countries around the world, this issue has conflict status, and would be a prime case for Internet mediation. It would empower Internet cryptography activists in putting their case, within the Internet community as well as without.

4 Software Patents and Copyright

Software patents applications being approved in increasing numbers in the United States pose a threat to the future development of the software arts within the Internet community. The interest of the United States companies benefitting from these patents appear to be outweighed by the wider interest of the Internet community.

Critics of US software patent laws say inventors of software are already protected by copyright and trade secret law, and inappropriate patents are approved, eg ìobviousî applications, which may already be well known by programmers or which may be reinvented over and over in the normal course of software development. They claim software developments, whether patented or not, build on prior art, and thus patents are inappropriate. Central to the debate, however, is the charge that large companies are viewing patents as an opportunity to profiteer at the expense of the international software development community.

The wider implications of this perceived threat are important. The proliferation of software patents in the United States may create a heavy international impact, on small businesses and community organisations seeking software to develop their operation.The running of, say, hospitals and schools could be affected in countries inside and outside the US, because of the actions of US patents lawyers. Progress in developing new networks and extending functions of the Internet would also be affected by the cost of extended patents searches and payment of royalties. Mediation may moderate debates such as the Unisys/Compuserve GIF row of 1995, in which a demand for royalties from Unisys led Compuserve to ask its software community to pay royalties on use of the Lempel Zev Welch Graphics Interchange Format. [Shearer, Vermeer, 1995]

After a threatened consumer boycott, Compuserve developed another GIF specification. If it is accepted that Internet users wish to encourage the best technology towards development of the Internet and global information systems generally, then the argument by large corporations that they have a right to make profits from software patents, (which may incidentally damage development), is seen as a claim with less merit. Presumably, it would then be the aim of Internet users to facilitate a global harmonisation scheme which would protect intellectual property rights, but not past the point where such rights hinder development of the software arts.

In the area of copyright the interest of the Internet community requires a balance to be struck between the need for authors to receive renumeration, the need for the development of comprehensive content bases, and the process of browsing ,without which the Internet cannot develop to full potential. The best outcome is likely to be a charging solution which will have the support and uptake of the Internet community. In the US, proposed legislation, the Information Infrastructure Copyright Act, has been criticised by the ACM public policy committee chaired by Barbara Simons :

ìOur concern with Section 106 (of the Act) stems from the definition of the word ìcopy.î Whenever a work in digital form is loaded to the RAM of a computer, or stored in the ìcacheî of a computer system, a ìcopyî has been made. The Bill as currently drafted could be interpreted to mean that the routine downloading into RAM or caching of a file could be considered infringement, even if this copying is not permanent and is incidental to the otherwise lawful use of the copyrighted material.î [Simons, 1996]

ACM copyright policy recognises a need for browsing of material to take place on the Internet. A further concern of the policy committee is tht copyright holders may, under the proposed legislation, be able to access information about users.

5 Pornography.

The issue of pornography has been a major creator of trouble for Internet users. Legislation passed in the US with the intention of limiting the spread of pornographic material , which would have limited free expression in many areas, has been countered by a finding by the US court of Appeals. Plaintiffs were individuals and organisations from the Internet community, which also organised a ìBlue Ribbonî

Campaign against online censorship. The Court found in June, 1996, that provisions of the Communications Decency Act were unconstitutional.

Threatened court action from a Munich official caused Compuserve to drop a number of listings from its international newsgroup delivery for a time in 1996. This is an area where consensus cannot be achieved about standards within countries, between countries, or within the Internet, despite exhaustive discussions on newsgroups such as alt. censorship. However, progress may be made, firstly, in accurately assessing the extent of the problem. Mainstream media reports have been shown to be inaccurate and misleading.The July 1995 issue of Time magazine fuelled the debate about the Bill. Timeís cover art depicted a horrified child looking into a computer screen, accompanied by the headline ìCyberporn. Exclusive: A new study shows how pervasive and wild it really is. Can we protect our kids - and free speech? [Time, 1995]

The major significance of the resulting furore on the Internet may be that the Internet community showed how it could potentially protect itself from the effects of irresponsible journalism. The Internet-based analysis of the story and its construction could be read as a critique of the operation of media sources formerly considered to have the highest professional and ethical standards. And in itself, it creates a defence for the right of Internet users to establish or defend the truth within their new public forum, the Internet. A WELL discussion, monitored by Brock Meeks [Meeks, 1996] included allegations that Time magazine:

- ignored efforts by a prominent academic to warn journalists that the study which featured in the story may have major failings.

- refused to release the study, which was conducted in an academic institution, for peer review.

- allowed figures from the study to be reported in a way which was not consistent with balance and fairness (primary journalistic values)

- allowed incorrect statements to be made about the extent of the study.

- oversensationalised the story by using ìshockî artwork.

It could thus be in the best interests of the Internet community to gather data on what is actually happening within the area of pornography, to assess:

- whether human rights abuses of subjects used for pornographic material are taking place

- whether the Internet is being used as a marketplace for the sale of pornography

- Whether pornography is interfering with the important "public forum " functions of the Internet, and whether it is limiting access by women and children to important information and public service resources. Working out definitions of Internet ìerotica,î ìpornography,î and ìsexual services,î could be a valuable mediation function in assisting the process of tagging or rating material for screening for protection of children.

Many women from the more developed countries, accord with the traditional feminist stance against pornography, and women of fundamentalist religious beliefs also disapprove of it. To them these values may overwhelm other considerations, such as protecting completely free speech. In discussing freedom of speech considerations, it thus becomes necessary for the Internet community to recognise the ìspecialî demographics of pornography as a political issue, and that sociological considerations may also be taken into account. If the Internet is to develop as a universal public forum, with the ìPublic Othereî characteristics identified by the German philospher Jurgen Habermas, [Habermas, 1989] it may be argued that women, as well as children, have a need to feel comfortable in the public spaces of the Internet, and that they should not be confronted with pornography in such ìpublic spaces.î While the Internet community has as a major identifiable ethic, complete freedom of speech, this ethic does not disallow the development of particular Internet environments which may be identified as appropriate for full public access.

Thus it is possible for the Internet community to support the right of the Internet community to uncensored databases in principle, while accepting that some interventions in this area may be necessary. These interventions may include the further development of screening software; self-policing of certain public forums, public identification and exposure of certain pornographic operations.

Mediation may be particularly effective in this area, in approaching the problem of pornography from two angles : that is, establishing the point at which adaptation of the Internet environment itself, or the use of screening and tagging of material , may constitute a defense against attacks by national interests, and by mediating directly with those outside interests to maintain the quality of free speech which is so valued by, and essential to the future of, the Internet community.

Conclusion

The idea of mediation in the Internet environment is a logical progression from recognition that the emerging Internet communications system faces serious political and regulatory problems. Though the idea of self-regulation appears to strike at the early Internet "dream" of complete freedom, mediation, rather then being seen as a mechanism of control, mays be viewed as a positive protector of Internet values. That is, it would allow the Internet community to draw a line beyond which it would oppose interference from national interests, in the interests of developing its own structure to its full social and technological potential. The mediation concept recognises social interactions developed by the Internet, by proposing solutions which are "social" in nature, that is, rely on community exposure, debate, and peer pressure. The framework favours education over coercion but adheres to certain Net specific principles, allowing interventions to take place. The proposal is constructed as a starting point of discussion and debate as to how freedoms valued by the Internet community may best be protected.

The problems of expensive, time-consuming and complex legal solutions might apper to be overcome by the employment of a mediation structure, (though legal solutions may be part of a resolution of conflict) but the mediation process remains imprecise. In the realms of evolving social and political structures within the Internet, that quality may be seen as advantageous. The immense technical and procedural difficulty of ìvote-takingî on issues is avoided. However the potential problem of ìcaptureî of a mediation group by a particular national group or ideology remains, and is likely to do so until the Internet community develops its own secure ideological base , for example a rational global ìmeta-ethic.î

References

[schwerin, 1995] Schwerin, E. (1995) Mediation, Citizen Empowerment, and Transformational Politics. Praeger Publishers US, pg 4.

[Schwerin, 1995] Schwerin, E. (1995) Mediation, Citizen Empowerment, and Transformational Politics. Praeger Publishers US, pg 19.

[Electronic Rights and Ethics, 1996] Electronic Rights and Ethics, <http://www.zip.com.au/~pete/ere.html>

[SanJose Mercury News 1996] San Jose Mercury News, (February 8, 1996) "Net Users Back Neo-Nazis' Rights." pg 1.

[Shearer, Vermeer, 1995] Shearer, J. Vermeer, A. (1995) Software Patents and the Internet: Lessons from the Compuserve/Unisys Graphics Interchange Format Case Study. Journal of Universal Computer Science, 1 (5), 308-315.

[Simons, 1996] Simons, B. (June, 1996) Letter from the USACM Public Policy Committee to Henry J. Hyde, Chairman, House Judiciary Committee,US House of Representatives <http://www.acm.org./usacm/copyright_letter_696.html>

[Shearer, Gutmann, 1996] Shearer, J. Gutmann, P. Government, Cryptography, and the Right to Privacy.

Journal of Universal Computer Science, 2 (3) 113-137.

[Elmer-Dewitt, 1995] Elmer-Dewitt, P. (July 10,1995) Cyberporn. Time International Magazine. No.7, pg 46.

[Meeks, 1995] Meeks, B. (July, 1995) Internet Commentary on WELL discussion of Time Cyberporn story.

[Habermas, 1989] Habermas, J. (1989) The Structural Transformation of the Public Sphere. An Inquiry into a Category of Bourgeois Society. Massachusetts Institute of Technology, US.

The MicroWeb Toolkit: Bringing the WWW to the Classroom

Judi R. Thomson and John E. Cooke and Jim E. Greer
ETHER Group
Department of Computer Science
University of Saskatchewan
57 Campus Drive
Saskatoon, Sk. S7N 5A9
CANADA
{thomson, cooke, greer} @cs.usask.ca

Abstract: Computer applications that facilitate the use of the World Wide Web within elementary and secondary education must provide support for educators to locate usable materials quickly and easily; they must minimize the amount of time spent waiting for files to traverse the network; and they must deal sensitively with censorship. The MicroWeb Toolkit is designed to provide software tools to organize and present educational information using World Wide Web resources. The MicroWeb Toolkit assists with the construction and organization of resource collections rather than with the authoring of individual pages by providing tools to organize and classify WWW resources around a specific topic. A MicroWeb collection is designed to minimize the cognitive overhead of the user and to reduce network traffic through resource caching.

1. Introduction

Computer applications that facilitate the use of the World Wide Web within elementary and secondary education must provide support for educators who need to locate usable materials quickly and easily; they must minimize the amount of time spent waiting for files to traverse the network; and they must deal sensitively with censorship. The application must also support the hypermedia nature of the Web, preserving structure whenever possible. The MicroWeb Toolkit facilitates the organization and classification of Web resources into topic-specific collections. Each collection consists of a series of concept pages and example pages. The concept pages are created by the collection developer using a concept map and the examples are resources found or placed on the World Wide Web. While using a collection, the learner is in control of pacing and browsing sequence but the educator has ultimate control of the material presented and how that material is linked to the concepts. When using the collection, learners view many different examples of concepts and gradually construct a personal understanding of the material.The process of creating concept maps, finding resources, and constructing the collection is time consuming, even with good tools. Fortunately the end result can be an extremely portable set of reference materials that can be shared widely. Although anyone, even children, could create MicroWeb collections, the MicroWeb Toolkit has been designed primarily to allow a "subject expert" to create collections for a particular topic and then supply those collections to teachers of that topic. Collections are usable on computer systems that can support a graphical Web browser. Teachers will not have to learn to use the MicroWeb Toolkit in order to benefit from the collections created by it because they can borrow entire collections, or parts of collections, from other educators.

2. Using Hypermedia Collections

The Web provides flexible access to unlimited amounts of material but, as with any form of hypermedia, the use of the Web is not without difficulty. For example, communications over the internet are frequently slow yet it is difficult to minimize the network traffic because there is no database management system for the Web. Automating resource updates is impossible because the location of copies of Web pages cannot be determined. The most frequently discussed problem that users of hypermedia resources encounter is a feeling of disorientation often referred to as being *lost in hyperspace* [Balasubramanian 1993]. Users who become lost need help to retain a sense of context about the material. The effort of maintaining this context while browsing a hypermedia resource has been called cognitive overhead [Thuring et al. 1995].Often users of the Web have difficulty finding useful material. A user must sift through much irrelevant material before finding the nuggets of information that match his or her interest. If the hyperlinks within the resource do not reveal to the user where the desired information is located, the resource lacks information coherence any may further impede the user's efforts to find information.To make the Web a usable tool for education, the issues of cognitive overhead and coherence must be addressed. Additionally, educational users of the Web must have methods to ensure that learners work with information that is suitable. Cognitive overhead can be reduced by providing navigational support as well as ensuring a consistent user interface. A coherent resource collection will allow the user to construct an accurate mental of the topic. The collection coherence can be increased by providing an overview of the hypermedia collection that shows the structure of the collection information [Thuring et al. 1995]. Any utility that gives the user an overview of the resource space surrounding the current position in the hypermedia resource can be classified as an overview diagram [Catledge and Pitkow 1995]. Overview diagrams can orient users to the surrounding hyperspace and to the content organization, affecting both cognitive overhead and coherence. In order to provide information about the hyperspace, the navigational tool must have access to information about hyperlinks. Usually this link information is stored in a database separate from the hypermedia system and the database is used to create the overview [Andrews et al. 1995; Roescheisen et al. 1995].Information about a hypermedia collection can also be communicated through annotations. An annotation is text or graphics added to a hypermedia resource with the intent of conveying extra information to the user. The most common form of annotation for World Wide Web hypermedia resources is the personal bookmark file, or hotlist. Annotations become more useful when they can be shared with other users such as through a trail or system of landmarks. Another method of sharing is to create and maintain a database of annotations provided by a particularly type of expert user. This database can then be accessed by users to find out what others have to say about a resource. The information in the annotation database could also be used to create tours and trails [Roescheisen et al., 1995; Andrews et al., 1995].An adaptive annotation system provides individualized information about a node to specific users [Brusilovsky, in press]. Adaptive annotation techniques could provide useful navigational aids for hypermedia collections by changing the annotations presented to suit the needs of the user. Users are presented with a stable view of the hypermedia application and are free to navigate the hyperspace in any fashion but navigational support is present at all times. Adaptive annotations can be something as simple as different colors for links or a complicated system that ranks resources in the collection based on resource position and content and annotates the collection based on the ranking [Tomek and Maurer, 1992].

3. The MicroWeb Toolkit

The MicroWeb toolkit is designed to provide software tools to organize and present information using World Wide Web resources. Many software packages are available for use by WWW information providers and most provide support only for the creation of individual Web pages. The MicroWeb toolkit assists with the construction and organization of resource collections rather than with the authoring of individual pages. The MicroWeb toolkit enables educators to present coherent hypermedia collections to learners using standard WWW browsers. The hypermedia collections, referred to as MicroWeb collections, can be presented either as a stand-alone hypermedia resource from which the learner may not stray, or as an application that allow connections to the larger World Wide Web.A learner uses a MicroWeb collection either by connecting to a special proxy HTTP server, or by loading collection documents from a local disk. From the initial collection document, the learner may select hyperlinks to navigate through the MicroWeb collection. The MicroWeb toolkit adds a header to each resource in the collection to give the learner extra information about the resource and the collection. The learner may, at the teacher's discretion, view Web pages that are not part of the MicroWeb collection. Clearly identified non-MicroWeb pages contain the standard MicroWeb header including a hyperlink back to the most recently viewed collection page. The learner may read through any resources in the collection, view pictures and animation sequences, and listen to sound recordings, all at an individual pace. The learner may make notes (annotations) about any of the resources which are stored for future reference. Annotations added to resources within the collection can assist the learner to navigate through the material. For learners who wish to find specific material, a table of contents is available for the collection in the form of a hierarchical concept map. The concept map allows the learner to see how sub-topics are related to one another as well as provides the learner with an overview of the entire concept. The learner gradually constructs his or her own understanding of the collection topic by reading, viewing, and listening to the material.

4. Creating a MicroWeb Collection

The components of the MicroWeb Toolkit are grouped into three modules. Two of the modules, the Resource Collection Proxy and the Collection Viewing Proxy function as proxy HTTP servers implemented using the STRAND toolkit [Brooks et al 1995]. The third module, the MicroWeb Control Environment allows the developer to organize and modify the collection. Educators who wish to create MicroWeb collections can benefit from a variety of components, including: a *Resource Collector*, a *Compiler/Interpreter*, an *Annotator*, a *Concept Mapper*, a *Learner Monitor*, an *Access Controller* and a *Collection Manager* [Figure 1]. Each component communicates with the MicroWeb database, which stores the MicroWeb representation of the resource collection. The information in the database includes meta-information about each WWW resource in the collection, links between collection resources, a representation of each of the concepts for the collection, and information about users of the collection.

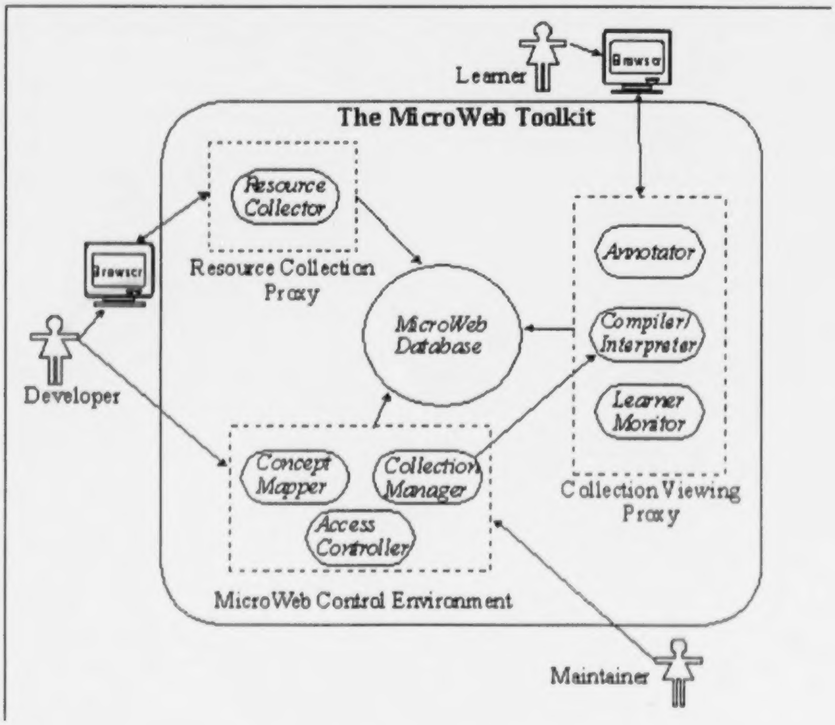

Learner
Browser
The MicroWeb Toolkit
Resource Collector
Resource Collection Proxy
Browser
Developer
MicroWeb Database
Annotator
Compiler/ Interpreter
Learner Monitor
Collection Viewing Proxy
Concept Mapper
Collection Manager
Access Controller
MicroWeb Control Environment
Maintainer

Figure 1 The MicroWeb Toolkit

4.1 Developer viewpoint

A MicroWeb collection cannot be created without resources and likewise cannot be created without a concept map to provide some organization to the collection. It is a matter of individual preference whether the first step is to find Web resources or to develop the concept map. In reality it is more likely that the developer will move between the two tasks until both are completed. The *Concept Mapper* allows the MicroWeb developer to gradually refine a diagram of the concepts to provide an abstract model of the general topic similar to the approach of [Zhao and Patel 1995]. The concept map for a new collection is begun by entering keywords to label the concepts to be included. The concepts are manipulated and moved around much as one would manipulate information within an outline for a document. Unlike typical outlining tools, the MicroWeb concept mapping tool allows concepts to be connected to other previously defined concepts within the outline. The result of the concept mapping process is a network-like description of the concepts for the MicroWeb collection. A graphical representation of MicroWeb concept maps becomes the primary interface between the developer and the concept mapping tool. At any point during concept mapping the developer may begin (or resume) collecting resources. Resources are simply Web pages or files that reside somewhere on the World Wide Web. They may be authored by the developer or by someone else. Resources are collected using a Web browser connected to the Resource Collection Proxy, which uses the *Resource Collector* component. The proxy server inserts an additional hyperlink into each resource to allow the addition of Web pages to the MicroWeb collection. Enough information about the resource is added to remind the developer of its contents and title. It is the responsibility of the developer to ensure that the original authors of the Web pages are amenable to having their resources included in the MicroWeb collection. The developer works alternately with the *Resource Collector* and the *Concept Mapper* to match the found resources to concepts in the map. The collected resources are linked into the MicroWeb collection using the concept map as a guideline for creating linkages. At first, each resource is linked only to relevant nodes in the concept. As the developer adds connections to the collection, linkages can be created between resources as well as between concepts in the database. For each resource the developer moves an icon representing the resource to the concept (or concepts) to which it corresponds. The linking process is very much like moving files around in a graphical file management program. As the map is adjusted by the developer, the *Concept Mapper* adjusts the database representation of the MicroWeb collection [Philip 1996]. Once the developer is satisfied that all of the available resources are linked to appropriate concepts, it is possible to ask the *Collection Manager* to analyze and check the consistency of the collection. The analysis detects concepts for which there are too many linked resources (indicating that some subdivision is likely desirable) or for which there are too few resources. It also checks that resources are sufficiently connected to the rest of the collection. When the developer is finished, the collection can be used. The developer can choose to create a collection that requires no internet access by using the *Compiler* or can create either a cached or direct-access collection using the *Interpreter*. A cached collection requires only intermittent access to the World Wide Web to update documents, while a direct-access collection always accesses the Web and retrieves the most recent version of documents. Finished collections can be viewed by connecting a Web browser to the Collection Viewing Proxy.

4.2 Teacher/Trainer Viewpoint

Although it is possible that the developer of a MicroWeb collection will also be the teacher using the collection, it is anticipated that frequently the developer and teacher will be two different people. Some of the features of the MicroWeb toolkit are designed specifically for the use of the teacher who uses MicroWeb collections for instructional purposes.It is essential that the teacher or collection designer be able to communicate additional information about the pages in the MicroWeb collection to the learners using the collection. Since the collections can be used independently, it is necessary for that communication to happen asynchronously. Annotations can provide an appropriate method for this kind of communication. After the collection is created, the teacher may use the *Annotator* component to add annotations to any of the resources in the collection through an HTML form similar to an email form. When an annotation is added, information about the person making the annotation, the MicroWeb resource, and the text of the annotation is stored as part of the database representation of the collection.Several factors determine how much control a teacher needs over learner access to the World Wide Web. Learners who are new to the Web environment will likely be more productive in their initial experiences if the hyperspace they can explore has definite boundaries. Similarly, learners who are not mature enough to judge the suitability of Web resources for their purposes can benefit from restricted access to the World Wide Web. Learners with more experience and maturity may benefit from some guidance, but likely would be able to manage working through the entire hyperspace of the WWW. The teacher uses the *Access Controller* to set access limitations for learners. Three levels of access are possible from a MicroWeb collection: denied, semi- restricted and unrestricted. Learners for whom outside access to the WWW is denied may view only those resources that are part of the MicroWeb collection. Learners with semi-restricted access to the Web are allowed to view other Web pages from selected portions of the Web. Semi-restricted learners may browse through particular resource hierarchies on specific Web sites but not jump to pages that are part of a different Web site. The site-based restriction allows the learner to thoroughly inspect a Web site that the educator has deemed appropriate without travelling too far into the Web. Unrestricted Web access allows users to view any non-collection WWW pages but the toolkit provides a hyperlink within the MicroWeb header allowing the user to quickly return to the MicroWeb collection. As learners work with a MicroWeb collection, records can be kept of the pages that a learner examines and the time spent on them. Teachers may view these records for evaluation or planning. The records allow teachers to retrace the path of individual learners through the MicroWeb. The teacher may define a list of required pages for specific learners to examine or recommended traversal paths. The MicroWeb toolkit can use the learner records and these lists to make suggestions about which pages the learner should look at. In this way learners can be guided into looking at specific portions of a MicroWeb. This feature is the focus of a related project [Philip 1996]. MicroWeb resources can be cached locally to reduce the need to connect to the Internet and the time spent waiting for resources to be retrieved. While this feature reduces the cost of connection and network traffic, it may present problems when the original author changes the resource. As resources in their original location are modified, the MicroWeb collection will become increasingly outdated. The *Collection Manager* component can automatically check for updated documents. When an update check is performed, the *Collection Manager* checks the original location of each resource in the collection to see if the original has been modified. The teacher is notified of any modified resources and can decide if the modified resource should replace the original in the collection.

5. Conclusions

The design of the *Resource Collector*, *Compiler/Interpreter*, *MicroWeb database* and *Concept Mapper* has been completed. Preliminary design is in place for the *Annotator* and *Access Controller*. Prototypes of the Resource Collection Proxy, Collection Viewing Proxy, and the MicroWeb database have been created and tested on a sample resource collection. The *Concept Mapper* and the *Learner Monitor* are under development as part of a related research project. It is intended that field testing of sample MicroWeb collections will begin soon, using a locally cached collection. The entire MicroWeb Toolkit will be tested for functionality upon completion of the components. Plans are underway to develop metrics that facilitate the evaluation of MicroWeb resource collections. We anticipate that the complete MicroWeb Toolkit will simplify the process of organizing WWW resource collections. It will provide tools for structuring concepts and organizing Web resources within that structure. The resulting collection should be coherent and have several different navigational tools to support learners. The toolkit allows educators to take advantage of the resources on the WWW while providing solutions to many of the problems that arise when using the Web. Learners will not become lost using a MicroWeb collection because navigational support is always present, yet they will gain valuable experience using hypermedia collections. Educators can use MicroWeb collections built by others, which can reduce both the teacher time spent finding the resources and the network time consumed by retrieving the resources. Finally, educators are assured that learners will remain on topic and view only appropriate material.

References

[Andrews et al. 1995] Andrews, K., Kappe, F., & Maurer, H. (1995). The Hyper-G Network Information System. Journal of Universal Computer Science, 1, 206-220.

[Balasubramanian 1993] Balasubramanian, V. (1993). Hypermedia Issues and Applications, A State of the Art Review. Independent Research Report as part of Ph.D. Program, Graduate School of Management, Rutgers University, December 1993.

[Brooks et al. 1995] Brooks, C., Mazer, M., Meeks, S., & Miller, J. (1995). Application Specific Proxy Servers as HTTP Stream Transducers. Fourth International World Wide Web Conference, 1995, Boston, MA, USA.

[Brusilovsky in press]Brusilovsky, P. (in press). Methods and techniques of adaptive hypermedia. User Modelling and User Adapted Interaction (to appear).

[Catledge and Pitkow 1995] Catledge, L., & Pitkow, J. (1995). Characterizing Browsing Strategies in the World-Wide Web. Journal of Computer Networks and ISDN Systems, 27.

[Philip 1996] Philip, T. (1996) The semantics of next in a hypermedia resource. MSc Thesis (in preparation), Department of Computer Science, University of Saskatchewan.

[Roescheisen et al. 1995] Roescheisen, M., Mogensen, C., & Winograd, T. (1995). Beyond Browsing: Shared Comments, SOAPS, Trails, and On-line Communities. Journal of Computer Networks and ISDN Systems, 27, 739-749.

[Thomson 1995] Thomson, J. (1995). Deep Sea Diving or 'What to do when you are tired of surfing?'. In T. Kusalik (Ed.), Proceedings of the Seventh Annual Graduate Symposium on Computational Science (pp. 17-28). University of

Saskatchewan: Department of Computational Science.

[Thuring et al. 1995] Thuring, M., Hannemann, J., & Haake, J. (1995). Hypermedia and Cognition: Designing for Comprehension. Communications of the ACM, 38(8), 57-66.

[Tomek and Maurer 1992] Tomek, I., & Maurer, H. (1992). Helping the User to Select a Link. Hypermedia, 4, 111- 122.

[Zhao and Patel 1995] Zhao, Z., & Patel, D. (1995). An Online Resource-based Learning Environment. In D. Jonassen & G. McCalla (Eds.), Proceedings of the International Conference on Computers in Education (pp. 315-323). Singapore: AACE.

A Web Site for counceling on students' future courses

Tsutomu Matsumoto*, Tomonori Wasaki*, Mitomo Okayama*,
Katsunori Horikawa**, Toshiriro Hayashi***, Mike Nga****

* Kumamoto National College of Tecnology ** The Intrigue Corporation
*** Faculty of Science and Engineering, Saga University
**** Omron Management Centre of AsiaPacific

E-mail : matumoto@ecsrv.ec.knct.ac.jp

Abstract

We have done the counseling on future courses to students using e-mail. Based on the experience we have designed the Web Site for couseling future courese. In this paper we describe the scheme of Web Site which has employment opprtunities and enrollment universities as a data base and how to collect URL which offer job opprtunities on the Internet. Furthermore we describe how we counsell the students using this system.

1. Introduction

The counseling on future courses is one of the most important non-teaching activities in Japananese College or University. Economic recession starting in 1990 has caused the shortage of employment opportunities. In this academic year, Apr. 1995-Mar. 1996, we have done the counseling on future courses to graduating students using e-mail. We provided employment information, vocational advice, information about enrolling to universities and discuss how to make recommendation letter[1]. It is much better for teachers in charge and graduates to access to the database on future course.

Nowadays the number of organizations or companies which offer the employment opportunities on the Internet is increasing. The issue then is that how to find them effectively on the Internet[2].

Though some Web sites are offered as search engines, with Web URL that could be collected manually or automatically[3,4]. It is not convenient for our students to find appropriate organizations and employment opportunity. They do not know which companies offer employment information on the Webs. In short they only want the Web that has URL of companies offering employment information. Furthermore they need various information about which company has a history of employing past graduates from our college, and the criteria to get the job and so on.

In order to improve these issues, we have developed a Web site with database for employment information and university enrollment information. This paper describes the design concept of our Web site. Further we report how to counsel future courses to graduating students using the Web site.

2. Design concept of database for counseling on future course

Graduating students of Technical Colleges in Japan get job offers before graduation. This is a Japanese tradition in getting a job. The traditional way is that staff from human resource(HR) department of companies visit teachers in charge. They ask teachers about possibilities of employment for graduating students as engineers. When teachers meet the HR staff of various companies, they try to get as much information as possible about criteria, vacancies, vacancies for women graduating students, etc. Another possibility is that HR of companies sent documentation related to employment by parcel service. We arrange these data and put them into the database.

Graduating students who want to enroll into universities look for universities accepting students from technical colleges. Many national universities send the documentation related to enrollment to Technical Colleges. We have to arrange them and we put them into the database.

The database that we should develop must offer two major information: one is employment information and the other is that Universities enrollment criteria. Of course employment information includes WWW URL that companies post want ads on the Internet. This database system should be a server-client system and also a WWW server, because we should be able to access it with any terminal.

Generally counseling for future course start on the beginning of each academic year. First we organize a special class for guidance. At the class we hand out brochures to students and explain outline of counseling and give various past information. This database should have like this brochure so that we do not need to print and hand it out. Figure 1 shows the scheme of database on above mentioned.

Database for employment information : Employment information that can be collected on the Internet are Web Page URLs and articles on News Group. Employment information directly come from organizations are arranged and stored into the database.

Database for university enrollment : Universities which publicize enrollment of students send guidelines for applications to Technical Colleges. Fundamental items, for example are exam date, exam subjects, the deadline for application, etc., are stored into the database. WWW of each University is linked from the database.

Brochure : The brochure of future course guidance is revised every year. This consists of text and table. We converted it to HTML file.

3. Database of employment

3.1 Net-search Web of employment

Net-search robots traverse all Web pages/Web sites which are constructed by Network Service Providers and companies. First, tagged HTML file of one Web site is transferred to our remote computer. At this point, only selective Web page's HTML files are transferred.

3.2 Analyzing HTML file

The analyzer builds a directed graph based on transferred HTML file. This graph explains structure of connected pages and connected sections inside one domain. Figure 2 shows one example. Then each page and section are scanned with breadth first search and are judged whether they are want ads. Judgment is done by including keywords. We choose some words: "employing graduating students", "salary" and "guidelines" written in Kanji(Chinese Characters) compounds. If it is a desired page or section, it is appended to the database. Appended data include URL and names of companies. The structure of this database is shown by Figure 3.

Generally there are two kinds of information: one is for recruitment of those already in the work force, and the other is for recruiting graduates. Even if this analyzer judges former category of employment, we allow it. Because it is still want ads.

3.2 Filtering articles on News Group

Sometimes new Web pages of want ads are announced. The articles are posted with job information. The tool of news filter is developed in order to select articles related to want ads. This tool finds articles and registers the URL and company name into database.

3.3 Employment information from organizations

In this section "employment information" means the guidelines for applicants which are directly announced by companies to us in the past and also in this academic year. We exported the past employment information that is transcribe to a card type database into relational database. Some companies send their organization's video tapes to us. The videos were converted to QuickTime files and stored them into the database.

3.4 Reports on interviews and examinations

Normally when Japanese companies employ graduating students, in order to evaluate applicants, HR staff interview and conduct examinations on major subjects and the liberal arts, for applicants. Students who have taken the examinations and interview for getting a job submit reports to us. This information is very important so we put it into the database.

4. Database of counseling university enrollment

4.1 Information of Universities

Students who want to enroll to University are interested in research activities and curriculum of each university. We collected Web page URLs of universities which admit graduating students of Technical College. The URLs are put on the clickable map of Japan.

4.2 Guidelines for applicants

Exam dates, exam subjects, which university requires a recommendation for students to apply, etc., are important for students who hope to enter university. These information are drawn from guide lines which came from university. And they are put into database as text data.

4.3 Past achievement

Grades, after school(extra-curricula) activities of past students who were admitted to the university gives much information to students who are keen to enroll. When any student choose a university, they will have access to much reference data. So we put the data into this database. Only administrator of this WWW database can access personal information. The aggregated data provide anonymous information of the contributors. With this method the database has been built.

5. Various documentation

We have handed out four kinds manuscripts to the students in the extra class for future course counseling. They are more than forty pages. Though we are willing to print and bind them every year. If the documents are on the WWW, we just need revise them. So we put the documents on the WWW.

6. Structure of Database

Figure 4 shows the structure of this system. A CGI generates SQL instruction set in order to execute user request, and SQL instruction set goes to relational database. Results influence to WWW browser, user's screen. We have built the database using BUTLER and TANGO which are commercialized by Everywhere Co., Canada.

7. User Interface

Navigation is via some icons on the Home Page, e.g.. Home page of "counseling for getting a job", Home Page of "counseling on the choice of University", Home Page of "various documentation for thinking your future", etc.,. Name of Home Page are displayed with Japanese-Kanji. Click on desired icon to go to desired Home Page. For example, If you want to search the companies which are located at Kumamoto, you type KUMAMOTO(in this case kanji-Japanese) into the dialogue box. Then CGI to do this command is executed, so CGI issues SQL instruction set to relational database. Results are displayed on the Home Page through Web server. See Figure 5. If displayed companies open their information on the Internet, clicking its name will allow user into the Home Page. When you search the university, same applies.

8. Conclusion

In this paper we propose a scheme of Web site for counseling on students' future courses. Our proposed scheme is designed on basis of our counseling experience using e-mail and counseling with traditional printed matter and face to face meeting. This Web site was built up not only as an example but aims to be of practical use. Further we believe we could more effectively counsel the students with e-mail and this Web site in next academic year. Net-search and news filtering keywords search are used for judgment without natural language processing. We are afraid that wrong Web pages and news get mixed up with right ones. We will develop effective methods of net-search and news filtering as a news project on natural language processing. User interface, page design is continuously being improved by results and feedback through practical usage.

Another possible extension to future project is to cater for employers and organization with appropriate information on student's career choice, personality and aptitude assessment to realized are more harmonic match.

[1]T. Matsumoto et al : A practical report of guiding students to get job with efficiency using e-mail and our future plan(J), JSISE Jan. 1996

[2]H. Nomura : Intelligent Information Access to Text on Information Highway, JIPSJ, Vol.37, No.1, pp1-pp9

[3]Ykoster, M. : Robots in the Web: Threat or Treat?
[4]Yahoo http://www.yahoo.com/, Altavista http://altavista.digital.com/

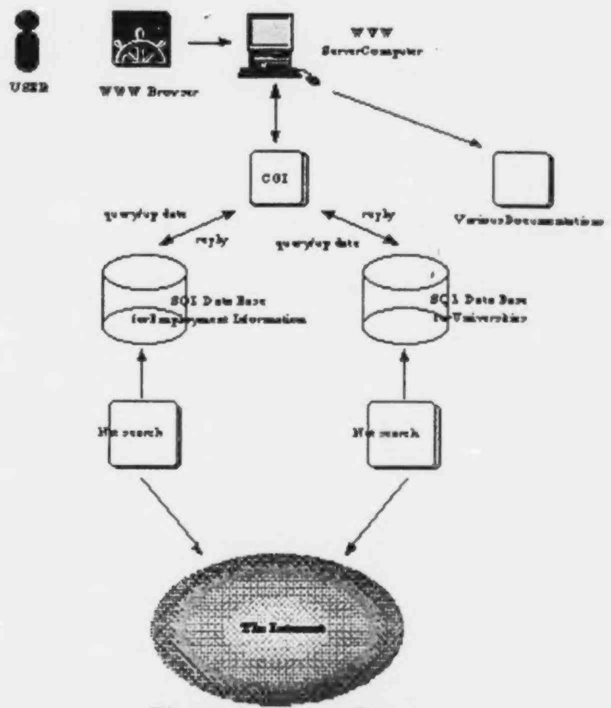

USER
WWW Browser
WWW
ServerComputer
CGI
query/up date
reply
reply
query/up date
VariousDocumentations
SQL Data Base
forEmployment Information
SQL Data Base
forUniversities
Net search
Net search
The Internet

Figure 1 System Structure

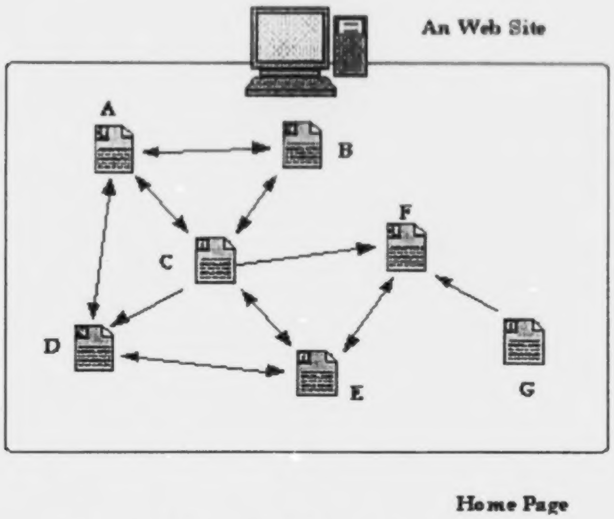

An Web Site
A
B
C
D
E
F
G
Home Page

Figure 2 Conected Page in one domain

BEST COPY AVAILABLE

URL	http://www.spin.ad.jp/cnavi/att/gis/recruit.html	TAB
TITLE	Recruiting Information	TAB
NAME	日本AT&T情報システム株式会社	TAB
DATE	1995/11/5	LF
URL	http://www.spin.ad.jp/tij/	TAB
TITLE	Welcome to TIJ	TAB
NAME	(c) Copyright 1995 Texas Instruments Japan Ltd. All rightsreserved.	TAB
DATE	1995/11/4	LF
URL	http://www.nri.co.jp/Personnel-Dept/index.html	TAB
TITLE	Cyber Personnel reception	TAB
NAME	Copyright (c)Nomura Research Institute, Ltd., All Rights Reserved.	TAB
DATE	1995/7/27	LF
URL	http://www.mew.co.jp/t35.html	TAB
TITLE	学生向けニュース	TAB
NAME	松下電工	TAB
DATE	1995/6/23	LF
URL	http://www.sumitomocorp.co.jp:80/student/index.html	TAB
TITLE	To Students	TAB
NAME	住友商事	TAB
DATE	1995/8/24	LF

Figure 3 Data Structure of Collcted URL on the Internet

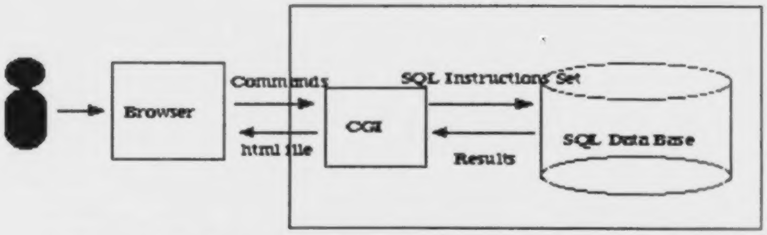

Browser
Commands
html file
CGI
SQL Instructions Set
Results
SQL Data Base

Figure 4 Relational among Browser, cgi and DataBase

Netscape: Search
Begins With
Reset Values

Figure 5(a) Searching Companies located at Kumamoto

BEST COPY AVAILABLE

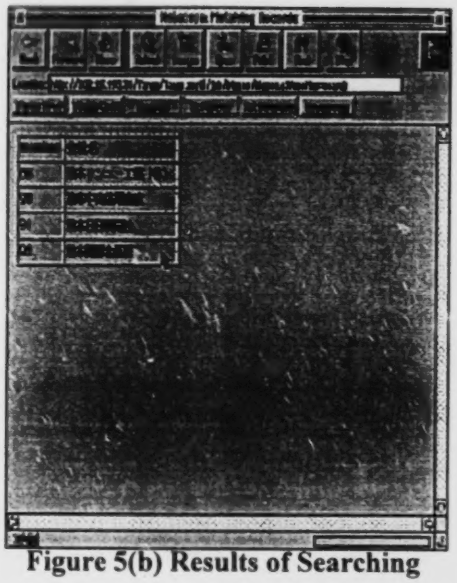


Figure 5(b) Results of Searching

BEST COPY AVAILABLE

ClassNet: Managing the Virtual Classroom

Mark J. Van Gorp
Durham Center
Iowa State University
U.S.A.
mvg@iastate.edu

Dr. Pete Boysen
Durham Center
Iowa State University
U.S.A.
pboysen@iastate.edu

Abstract: Education continues to move on-line through the World Wide Web. Classrooms of students and teachers are no longer restricted by time or distance. ClassNet (http://classnet.cc.iastate.edu/) is a tool which manages these virtual classrooms: It automates many of the administrative tasks associated with global Internet classes. Through a simple interface of Web forms, students can perform activities such as class registration, assignment or test submission, and grade retrieval. Meanwhile, instructor activities include managing assignments, controlling class enrollment, communicating with students, and monitoring student progress. This article highlights features of ClassNet's design and functionality and provides examples of its use.

Introduction

Virtual classrooms, virtual degrees, and virtual universities are actually explicit realities in today's educational community. Authors such as [Barker 1994] note that we are now educating geographically diverse populations and that "virtual [classrooms] will become as commonplace in higher education as the chalkboard once was" [p. 159]. One only needs to visit either the World Lecture Hall at the University of Texas at Austin (http://www.utexas.edu/world/instruction/index.html) or the Open University in England (http://www.open.ac.uk) to find concrete examples of this educational phenomenon [Shotsberger 1996]. Further, well designed activities via the Web can offer students a valuable learner-centered education [Stout & Thompson 1995].

However, how are these distance-oriented classrooms managed? Who or what handles details such as the registration, grades, assignments, portfolios, and tests? Some have used e-mail for handling a few of these details [Pitt 1996; Poling 1994; Wei He & Knapp 1995]. Others have programmers developing CGI-Scripts specifically for their classes [Dix, Allendoerfer, Jones, Lacey & Laurenzi 1995]. These scripts handle the student information (originating from a browser interface) and may or may not store the information centrally. Unfortunately, not all instructors involved in distance education have access to a programmer, some CGI-scripts already developed are too specific for use across a wide array of classes, and e-mail presents disk space and organizational problems [Wei He & Knapp 1995].

ClassNet (http://classnet.cc.iastate.edu/) is a general solution to this problem: Its purpose is to bring automated administrative functionality to global Internet classes. Classrooms and their management details are handled by a tool that organizes the diverse information needed to administer these classes. Students and teachers interact with this tool through a simple Web interface and thus only need access to a browser of their choice. Although ClassNet's purpose is not unique (http://west.ucd.ie/), many aspects of its design and functionality are.

ClassNet Design, Underlying Objects, and Their Functionality

ClassNet is a CGI-Script running underneath a UNIX HTTP server. The alpha release (January, 1996) is written in C++, but the future beta release (scheduled for August, 1996) will consist of Perl5 modules. Users interact with ClassNet through simple HTML forms which are not browser specific. ClassNet, in turn, acts as the gateway between a database of registered classes and the users.

ClassNet consists primarily of four objects: requests, classes, assignments, and class members. Each object provides necessary functionality for ClassNet's operation. Considering the myriad of responsibilities in classroom management (e.g. student drops and adds, assignment due dates, assignment corrections, grades, security, etc.), each object becomes rather complex in its own unique way. Abstract descriptions and portions of their responsibilities are given below.

Requests and Classes

The request object is the first object created when the HTTP server calls ClassNet. Its responsibility is to create an array of name/value pairs contained in the incoming HTML form. Forms are generally required to have four field names for identification purposes: username, password, class, and page name. The page name contains a value which identifies the user's request to ClassNet.

Essential duties of a class object are to check for class membership, to permit student registration, and to create and remove classes from the database. Anyone in the world can enroll for a class registered with ClassNet. Consequently, the class object keeps track of students requesting enrollment and generates an HTML form consisting of checkboxes and student names when requested by the instructor. All checked students are then enrolled in the class while any spurious requests are removed. Thus, the instructor has complete control of the enrollment process merely through a point and click interface.

Assignments

Assignments (tests and quizzes included) are HTML forms created by the instructor and are handled by the assignment object in ClassNet. These forms are stored in the instructor's Web directory and thus are not stored in ClassNet's database. To aid the instructor in the assignment construction process, an interactive cut and paste mechanism is currently employed: An instructor enters a question through a ClassNet form, and ClassNet sends back the HTML version of that question which can then be pasted into a growing HTML document on the instructor's end. This mechanism is being updated for the beta release of ClassNet: Instructors will be allowed to construct multiple versions of the same question, and they will be sent the entire form upon completion instead of one question at a time. Multiple versions of the same question will allow for automatic random generation of test/quiz questions to each student in an entire class.

Three question types are recognized by an assignment object -- multiple-choice, short-answer, and essay. Fill-in-the-bank questions are a type of short-answer, and true/false questions are a type of multiple-choice. Automated grading is permitted for multiple-choice and short-answer questions as both the instructor and student answers are stored in ClassNet's database. Currently, tests containing essay questions are e-mailed to a designated recipient for further grading and comments on the essays. Essay answers are not stored in the database but will be in the beta release.

When the instructor submits an assignment form, ClassNet recognizes this as an answer key. Subsequently, another form is returned to the instructor requesting information such as the due-date, the permission for students to view the answer key after the due-date, the partial and total points for each question, the answer range for numeric short-answer questions, and the requirements for grading short-answer string answers (e.g. enforce capitalization or spelling). This information is compactly stored along with the answer key in the database.

Because assignment answers for both the teacher and the student are stored in the database, students can receive immediate feedback on how they performed, while instructors can easily view how individual students or an entire class are doing. Students asking ClassNet for a particular assignment will receive both their answers and the instructor's -- provided the due-date is past and the instructor has told

ClassNet that students may view the answers. Instructors may also enter grades for students on other assessed in-class activities. Thus, ClassNet is also an on-line grade book which students can access from any geographic location. Although the instructor has access to all students' assignments and grades, student usernames and passwords permit students to only view their own.

Class Members -- Instructors, Students, Proctors

Instructor objects have the least restrictions and the widest range of functionality in ClassNet. For example, instructor activities may include:

_ registering classes (with approval from ClassNet administrators);

_ approving student-initiated registrations;

_ adding, editing and deleting any class members;

_ adding, editing and deleting assignments and grades;

_ creating assignments;

_ communicating with students; and

_ viewing student grade reports or assignments.

The second, and most frequently used of the class member objects, are students. Students can register for a class, submit assignments for grading, see their grades, and view their own graded assignments and answers. The ability to view assignments and grades on-line will alleviate traffic to the instructor or teaching assistant. Student requests for class registration must be approved by the instructor.

Proctor objects allow individuals designated by the instructor to verify that the actual student submitted his or her quiz. Proctors receive a list (HTML form of checkboxes) of all students who submitted a particular quiz or test and check those students whom they monitored. Subsequently, the quiz or test of checked students are graded and possibly e-mailed for further grading of essay questions while unchecked student assignments are left ungraded. Answers to essay questions in the beta release will likely remain in the ClassNet database, and the instructor will interact directly with the database to grade those questions. An e-mailed version will be available as well.

Organization of the ClassNet Database

ClassNet's database is currently organized as UNIX directories and file structures. All classes are stored on a secure disk accessible only by ClassNet administrators, and backups are performed daily. A typical class layout is:

classname

/ | \

admin ASSIGNMENTS students

/ \ / .... \

members REQUESTS student_1 student_n

/ \ / \ / \

STUDENTS TEACHERS GRADED UNGRADED GRADED UNGRADED

The bottom rows of directories, denoted by capital letters, contain the files needed for class management. The student and teacher directories contain individual class member files which hold five pieces of information: first name, last name, username, password, and e-mail address. The request directory contains those students requesting enrollment, and a student file is moved from this directory to the student directory when an instructor approves enrollment of that particular student. For security purposes, the request directory also contains an enrollment monitor that places a limit on the number of outstanding enrollment requests. The assignment directory contains all the HTML assignments for a class. These files hold only the name, answer, and grading specifications for each question. Security plays a role here as well: An incoming student assignment is checked against these files to see if it is indeed a valid assignment containing valid questions. An invalid incoming assignment is discarded and the student is notified. Lastly, the graded and ungraded directories contain each student's submitted graded and ungraded assignments.

Current Uses

Dr. Doug Yarger at Iowa State University has used ClassNet to manage daily weather forecasts of 250 students in a weather forecasting contest. Dr. Yarger's primary assignments are actually "forecast" forms which each student submits approximately 60 times per semester. He also uses ClassNet as an on-line grade book for additional in-class quizzes.

For the weather forecasting contest, students extract weather information from on-line products such as surface maps and satellite images and then submit forecasts for the following day. A forecast form asks students for 6 a.m. and noon temperatures, wind speed, wind direction, and precipitation, as well as explanations for their predictions. Because the forecast name contains a ClassNet reserved word delimited by @ symbols (e.g. DesMoines@NEXT_DATE@), it will be modified to a string representing the next day's date (e.g. DesMoines3_30_96). Consequently, one assignment form may be used for multiple assignments, and the students are bound to an explicit deadline of midnight.

Because Dr. Yarger's form consists only of multiple-choice and short-answer questions, ClassNet grades the entire student forecast automatically. If it contained essay questions, further human evaluation would be needed. When a student submits a forecast, ClassNet stores the student answers (provided the form has passed the validity check) and then grades all previously ungraded forecasts. It cannot grade the current forecast as the true answers cannot be identified until at least the next day. One of Dr. Yarger's teaching assistants later uses an identical form to enter the actual weather data obtained for a particular day.

ClassNet also allows a student to view any previously graded forecasts or assignments. A student first uses a ClassNet form to enter her or his username, password, class name, and request. If the request is to view assignments, ClassNet then sends back another form of radio buttons where he or she can select which assignment to view. After the student selects an assignment (or forecast) and re-submits the form, ClassNet responds by sending back an HTML document containing her or his answers for each question, the instructor's answer for each question, the points received and total points for each question, as well as the total points received and percentage correct for the assignment. If an assignment such as an in-class quiz were requested, ClassNet then sends back only a summary of the percentage correct and points received.

Another professor, Dr. William Gutowski, has been experimenting with ClassNet for on-line assignments. He constructs his own assignment forms through ClassNet's form construction mechanism.

Lastly, Mary Herring, a doctoral student and temporary instructor at Iowa State, is using ClassNet for a series of surveys. Her initial surveys consist entirely of essay questions, and her respondents subsequently answer these questions and submit the forms. ClassNet takes in the forms, formats the answers, and routes the answers on to the e-mail address designated by Mary. Her final surveys will be multiple-choice forms (Likert scale), and she may wish to either have these forms e-mailed to her or stored in ClassNet. In the beta release of ClassNet, we hope to incorporate item analysis which will help in analyzing or summarizing multiple-choice and short-answer surveys. Using forms in survey research answers some disadvantages listed by [Thach 1995] concerning the use of electronic mail surveys:

Forms through ClassNet can be set up to guarantee anonymity, and the layout of a Web form is generally much more inviting than an e-mailed questionnaire.

Features and Futures

ClassNet contains many features -- some of which were previously mentioned. These features, beneficial both to education and research, include:

_ global and device independent access where users are restricted neither to a geographic location nor to a particular hardware architecture or operating system;

_ easy classroom management integration with existing Web materials;

_ course-independent structure;

_ separation of course content from course management;

_ automated management helping instructors focus on teaching, not grading;

_ securely stored classroom management information (e.g. grades and assignments);

_ centralized accessibility to organized classroom data;

_ immediate and private access to individual assignments and grades;

_ capability to edit student grades and assignments or to automatically re-grade entire class assignments;

_ easy distribution and collection of surveys;

_ survey anonymity; and

_ free distribution.

Separation of course content from course management is a feature deserving special attention. This design decision was prompted by storage and efficiency concerns. Because ClassNet does not store Web pages, more storage is reserved for class registration. Additionally, instructors have total control of where their class content is stored and how it is delivered. ClassNet's efficiency is also enhanced: The server calling ClassNet must only respond to classroom management requests rather than both to management and content requests. A server handling requests such as viewing assignments, accessing and assigning grades, grading incoming and previous assignments, and communication with class members is a load in itself. Serving HTML pages of class content will burden the server further and unfortunately increase the "wait-time" of users. The management load presented by Dr. Yarger's one meteorology class of 250 students and additional teaching assistants has shown this to be an important design decision.

ClassNet is receiving a growing "wish list" for the future. Some of these requests are for additional options, while others are for new ways to improve existing options. Requests have included needs to:

_ provide additional support for inter-class communication;

_ store links to student Web page portfolios;

_ generate randomized test questions;

_ statistically analyze individual survey and test items;

_ calculate weighted grades and permit deletion of lowest scores;

_ store or route incoming data from Java simulations; and

_ further develop the assignment creation process.

These wishes are just a subset of a dynamically growing list: The ideas for future development are many.

Conclusion

The development of ClassNet has been an exciting and ongoing process of thinking how to improve education using the Internet. The use of the Web and CGI-Scripts have allowed distance education to become much more interactive. For example, the weather forecasting contest has recently been extended to other elementary, secondary, and post-secondary students and classes. This was not possible without the use of a tool such as ClassNet and required no modification of ClassNet's design. Our hope is that ClassNet will aid teachers in managing global or local classrooms as well as facilitate the learning process.

References

[Barker 1994] Barker, D.I. (1994). A technological revolution in higher education. Journal of Educational Technology Systems, 23 (2), 155-168.

[Dix, Allendoerfer, Jones, Lacey and Laurenzi 1995] Dix, J.A., Allendoerfer, R.D., Jones, W.E., Lacey, R.A., & Laurenzi, B.J. (1995). An electronic curriculum for introductory chemistry. Journal of Educational Technology Systems, 24 (2), 151-157.

[Pitt 1996] Pitt, M. (1996). The use of electronic mail in undergraduate teaching. British Journal of Educational Technology, 27 (1), 45-50.

[Poling 1994] Poling, D.J. (1994). E-Mail as an effective teaching supplement. Educational Technology, 34 (5), 53-55.

[Shotsberger 1996] Shotsberger, P.G. (1996). Instructional uses of the world wide web: Exemplars and precautions. Educational Technology, 36 (2), 47-50.

[Stout and Thompson 1995] Stout, M. W., & Thompson, J. L. (1995). Instructional design issues and the world wide web. Educators Tech Exchange, 3 (1) 24-35.

[Thach 1995] Thach, L. (1995). Using electronic mail to conduct survey research. Educational Technology, 35 (2), 27-31.

[Wei He and Knapp 1995] Wei He, P., & Knapp, S.D. (1995). Electronic reserve with WWW: A promising way to enhance classroom instruction. Journal of Educational Technology Systems, 24 (2), 119-125.

Modelling Alter Egos in Cyberspace: Who is responsible?

R.P. van de Riet and J.F.M. Burg [1]
Vrije Universiteit
de Boelelaan 1081a
Amsterdam, The Netherlands
Tel: +31 20 4447757
Fax: +31 20 4447653
{vdriet,jfmburg}@cs.vu.nl

Abstract: Alter Egos represent people in Cyberspace. An Alter Ego is a kind of intelligent agent who is active in performing actions in behalf of the person it represents. How these Alter Egos can be modelled and constructed is discussed. In this context the question whether Alter Egos can be held responsible is studied.

Introduction

Modelling in Cyberspace can be done from two different perspectives: one is modelling the information which is available, the other one is modelling the people living in it. With this we mean that people will be (are) represented by active agents, objects residing in computer systems. They represent real people when real people leave doing their business to them, as already happens with ordering goods and automatic payments. We call these objects Alter Egos. The approach presented here is a puretechnical one, in contrast with for example Erickson [Erickson 96] who describes the World-Wide Web (W3) as a social hypertext, in which the W3-nodes are becoming (social) representations of people.

We are studying how Alter Egos can be modelled when they are being used for home banking and aspects of privacy and security are involved. Also we came upon the question of what is meant when we say that these objects are responsible while developing certain modelling tools. These tools are being used to model complex behaviour of people and computer systems, communicating with each other. Dealing with people involves also dealing with responsible people. That means they can be held responsible for their deeds; they can be punished when they do not do what they promised. So communicating with people means that one always has to take into account that these people do not do what they promised (or committed themselves to), on the other hand computer systems always do what they are programmed to do (we do not talk about failures of these systems, but assume that they function correctly, according to the specifications).

Naturally, then the question comes up when people are represented by Alter Egos whether these Alter Egos can be held responsible, and what this means.

An Alter Ego for a Cybernaut

Cyberspace is considered to be the space in which people and information is connected through computer networks. Nowadays we see the effect of Internet; in the near future the new information superhighway will provide the facilities. In this paper the people in Cyberspace will be called *cybernauts*. One can look at them from two different perspectives: as active persons communicating with each other and with information systems (e.g. through the W3) and as objects who are themselves being dealt with, e.g. by governmental agencies or by Banks. The idea is that the Social Security Number (SSN) will be replaced by an object. Probably recorded in object-oriented databases in the same way as people nowadays are represented in ordinary databases using SSN.

---

[1] Supported by the Foundation for Computer Science in the Netherlands (SION) with financial support from the Dutch Organization for Scientific Research (NWO), project 612-123-309.

In [vdRiet 95], [vdRiet 96] we have introduced the notion of Alter Ego which combines both aspects: an object as *subject*, acting in behalf of its owner as a so-called intelligent agent, and as an *object* representing its owner.

The issue of responsibility is relevant for both aspects mentioned. However, we will be interested only in modelling the behaviour of an Alter Ego.

In Western Society citizens are represented by their SSN. It is not very far fetched to assume that in the near future these citizens will be represented by an object (in the object-oriented (OO) sense), where the SSN is replaced by an object identifier (or is the object identifier itself).

By using OO methodology it is possible to construct logically quite sophisticated objects representing these cybernauts in Cyberspace.

Consider first the possibility that a central agency, like the Civil Administration (CA) keeps track of all the personal properties of a cybernaut, such as: name, address, birthdate and marital status.

Consider next that a cybernaut can inscribe in a University U as student and has additional properties such as field, and year.

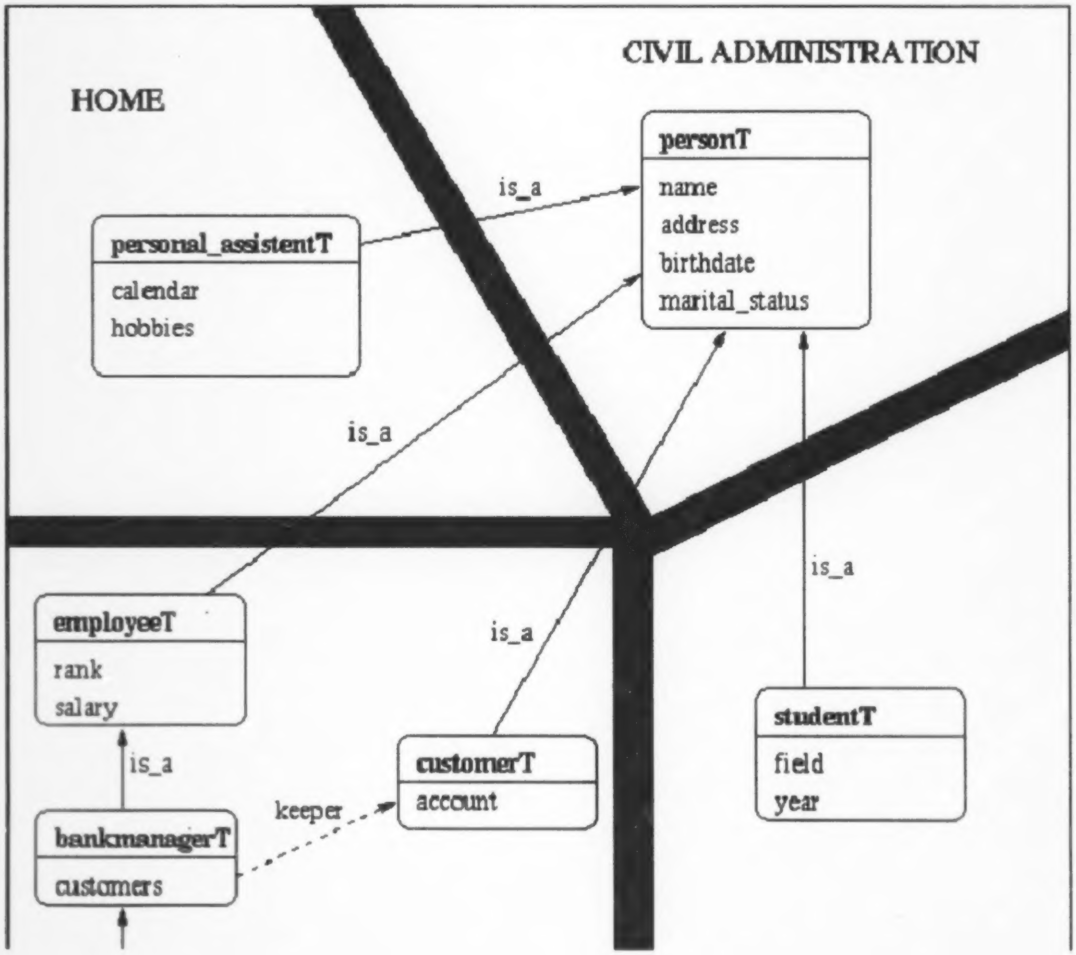

CIVIL ADMINISTRATION
HOME
personT
name
address
birthdate
marital_status
is_a
personal_assistentT
calendar
hobbies
is_a
is_a
employeeT
rank
salary
is_a
is_a
customerT
account
keeper
bankmanagerT
customers
studentT
field
year

Figure 1. Physical worlds of objects

Let us further suppose that the cybernaut may also be a customer of some bank B, having an account with a number and an amount.We can now define in Mokum (see [Dehne 94]) the types as they are given in Fig. 1. Evidently, we assume here that each cybernaut has a personal assistant object at her disposal, which may actually reside in a home-based PC or a personal digital assistant (PDA) a smart card or even a wrist watch.

Objects have (one or more) types, which can be structured in the form of an is_a hierarchy involving the inheritance of properties.

A cybernaut is always a person, so there must be an instance of personT, representing this cybernaut. If she is also a student at U and a customer of B there must also be instances of the types: studentT and customerT. These objects together form one object called the Alter Ego. Statically the structure of an object is quite simple. It is their dynamic behaviour which makes them very powerful and fit for their service as Alter Ego.

To each type a so-called script can be added in which the behaviour can be defined. In principle each object is a finite state automaton, which reacts on signals coming from elsewhere (usually messages sent by other objects to this object) in a predefined way. It is the way these scripts are defined that is studied in this paper.

Consider now an Alter Ego which is an object of type personT and of type studentT. We shall say that in this case there are two sub-objects: a person object and a student object; the Alter Ego object is the single object consisting of both sub-objects.

Logically these objects belong to each other, through the is_a relationship. Physically they are located in quite different worlds, with their own protection rules.

In Fig. 1 these worlds are shown. There are fire-proof walls between these worlds to protect private information. We will assume that private information is kept in attributes declared 'private' in Mokum terminology.

We will say a few words here about problems connected to the protection problems. In a more general setting these are dealt with in [Varadharajan 95], [Olivier 95], and in the more specific Mokum environment in [vdRiet 94], [vdRiet 95], [vdRiet 96]. We simply assume that the following rules from Mokum are implemented:

- the epistemic rule which says that private attributes of some type can be read and changed within the script of a subtype; the same holds for so-called keepers of a collection of objects of some type;
- the ontological rule which says that only the object itself or the keeper of a collection in which that object lies can read and change the attributes of that object.

Introducing the notion of collection and a so-called keeper of a collection, it is possible to define rules for maintaining integrity and security in a very natural way, almost as is done in actual practice, where a person is held responsible for keeping some rules. In our case an object can take the role of a keeper of some rules for a collection of (other) objects. Such objects can be the hobbies of some cybernaut (as we saw in the example type definition above of 'personal_assistantT'), but they can also be a collection of customer objects being kept by a bank manager.

We have thus defined within a bank organization how a bank manager can have an assistant who is *responsible* for a collection of customers but whose *responsibility* does not go beyond a certain limit (evidently, the Barings Bank did not restrict Nick Leeson to such a limit, which they painfully regret).

Above we used a few times the word *responsible* in a very natural way. The word means simply what it says: a certain task which has to be carried out, such as dealing with financial accounts of cybernauts.

Modelling Cyberspace using Color-X

Consider the situation of the University library borrowing a book to a student. We are working on a tool, called Color-X [Burg 95a, 95b, 95c], by means of which this situation can be described:

- very concisely, by means of a diagram,
- very succinctly using linguistic tools, such as WordNet, and
- very helpfully, because Mokum scripts for the objects involved can be generated almost automatically.

The diagram specifying the library situation is shown in Fig.2. It describes the behaviour of the library system and a human being: the user. After obtaining a pass the user is permitted to borrow a book, which she then has to return within three weeks. Later on we will change the example and put a personal assistant in between. The boxes in the diagram denote actions with some modality:

- **PERMIT** to indicate that the actions are allowed;
- **NEC** to indicate that the actions are necessarily to be performed by the library system. In case the actions are not performed, the library system has a serious problem: it is in an inconsistent state (comparable with the situation of a personnel database having a record of some person with a negative age).
- **MUST** to indicate that the actions have to be done, but there is a possibility that they are not done, within the prescribed amount of time; in that case there is a special arrow, the lightning arrow, which indicates what action is performed by the library system.We are dealing here with certain actions, such as returning the borrowed book, which the user should do, but she may decide not to do them, or simply she can be in the situation not being able to do them. We shall say that the library system has to deal with a *violation* in such a case.

We see here that MUST is closely related to the word *responsible* as MUST takes into account promises, not fulfilling promises and punishment.

Event Model for the Library System

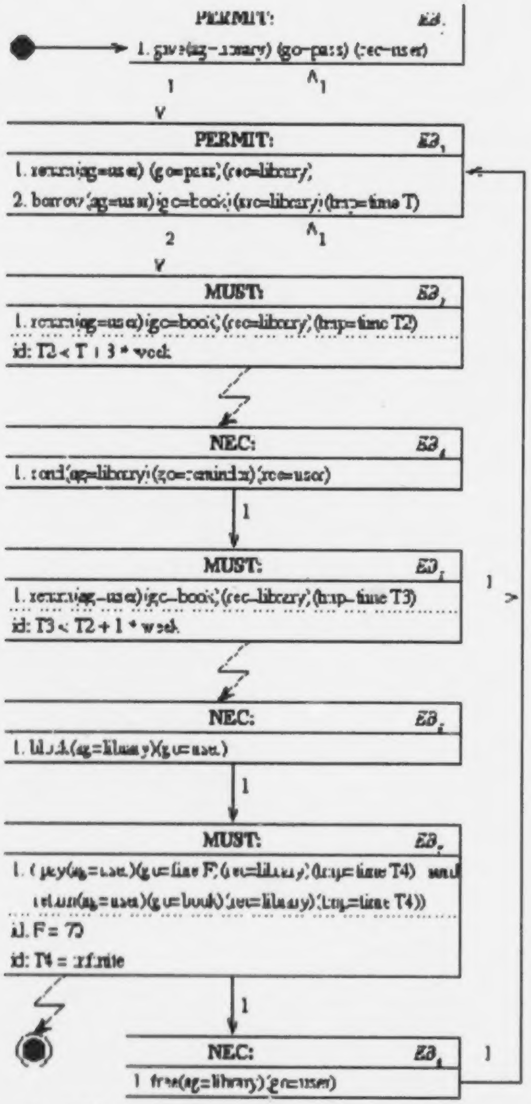

PERMIT:
1. give(ag=library) (go=pass) (rec=user)
1
1
PERMIT:
1. return(ag=user) (go=pass) (rec=library)
2. borrow(ag=user)(go=book)(src=library)(tmp=time T)
2
1
MUST:
1. return(ag=user)(go=book)(rec=library)(tmp=time T2)
id: T2 < T + 3 * week
NEC:
1. send(ag=library)(go=reminder)(rec=user)
1
MUST:
1. return(ag=user)(go=book)(rec=library)(tmp=time T3)
id: T3 < T2 + 1 * week
1
NEC:
1. block(ag=library)(go=user)
1
MUST:
1. (pay(ag=user)(go=fine F)(rec=library)(tmp=time T4) and
return(ag=user)(go=book)(rec=library)(tmp=time T4))
id: F = 70
id: T4 = infinite
1
NEC:
1. free(ag=library)(go=user)
1

Figure 2. COLOR-X Event Model for the library system

BEST COPY AVAILABLE

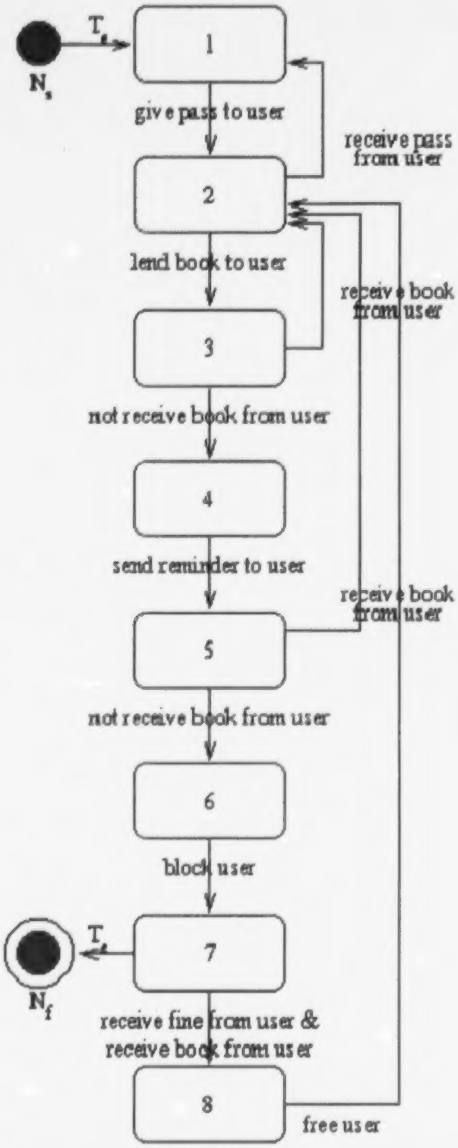

T_s
N_s
1
give pass to user
receive pass from user
2
lend book to user
receive book from user
3
not receive book from user
4
send reminder to user
receive book from user
5
not receive book from user
6
block user
T_f
N_f
7
receive fine from user & receive book from user
8
free user

Figure 3. State Transition Diagram for the library system

It is evident that being in an inconsistent state is much more serious than dealing with a violation. In the first case the system cannot be trusted anymore and serious measures have to be taken immediately. In the latter case it is a matter of duely waiting, because, according to the diagram, eventually, there comes a moment that the violation will be resolved. Either by the action of the user to return the book, or by the action of the library system to block the user, after several warnings. The nice thing about the diagram is that indeed such a situation will occur, because there are no MUST boxes without a lightning arrow leading to some other box.

The formulae written in the boxes are so-called CPL formulae; the syntax is very close to Functional Grammar (see [Dik 89]). Indeed, one can see what kind of things (objects) may occur as roles/parameters within the formulae: in the borrowing action, the user is the active agent, the library is the source and the book is the goal. In this way one can see which roles are being played by which objects.

In the actual Color-X tool heavy use is being made of the WordNet [Miller95] system, which is a lexicon consisting of the meaning (in a computer-friendly form) of some 90.000 words/concepts. Using this lexicon it is for example impossible to define an action in which the book is the actor of the borrow action. Also is it possible to generate finite state automata for the objects involved in our activities: library and user. The result is shown in Fig. 3. Note that what is "*to borrow*" for the user is "*to lend*" for the library. Such knowledge is taken from WordNet. Also, "*to give*" a pass for the library is "*to get*" a pass for the user.

In Color-X it is also possible to generate natural language sentences describing the model, as is illustrated in Fig. 4.

[if,a,library,gave,a,pass,to,a,user,then,

the,user,is permitted to,return,the,pass,to,the,library,or,

the,user,is permitted to,borrow,a,book,from,the,library]

[if,an,user,borrowed,one or more,books,from,a,library,at,time,T,

then,an,user,will have to,return,one or more,books,to,the,library,

at,time,T2, where,T2,is less or equal,T + 3 * week]

[if,an,user,borrowed,a,book,from,a,library,at,time,T,and,

an,user,did not,return,the,book,to,the,library,at,time,T2,where,

T2,is smaller than,T + 3*week,then,

the,library,is obliged to,send,a,reminder,to,the,user] Figure 4. Generated Natural Language sentences for the library system

Finally, it is possible to generate the script for the library system, acting as an object who is *responsible* for the bookkeeping.

What is responsibility?

From the Introduction we have: responsibility involves the characteristics:

a) free to keep or not to keep promises made

b) able to explain ones deeds

c) in case of fault: penalties

Let us give a few examples where the word *responsible* is being used:

1. In a CORBA document [OMG 91], p. 41, functions are defined for which the object adapters are *responsible*.

2. In the section above about Alter Egos, where we dealt with objects being keepers of collections of other objects, for whose integrity they are held *responsible*. 3. In the section about modelling: there is a user who is *responsible* for returning the book, but very easily one can replace this user/cybernaut by the Alter Ego it represents. Can we say that the Alter Ego can be held responsible for returning the book?

Let us analyze the three situations and compare them with the definition of responsible. For examples 1 and 2 it is evident that item b) can be realised: indeed, it is possible to program the adapter, the keeper and the Alter Ego in such a way that they can explain why they did what they did.

For items a) and c) it is however rather difficult to imagine that a piece of software such as an adapter, keeper or Alter Ego can comply to them. Either they are programmed (by a human programmer, who probably is the real one who is responsible) to act according to the promise, or commitment, which has been made, or not, but in the latter case they are programmed badly. We assume that that is not the case. A middle position could be that on purpose once in a while the adapter is not performing the function properly; in that case the programmer has not done the work properly (and should be punished). Also, what does it mean that a piece of software should be punished?

So, we come to the conclusion that principally software cannot be held responsible, so Alter Egos cannot be held responsible. That means that it is not necessary in the world of Cyberspace to take into account the MUST modality, when Alter Egos are being modelled. We shall see that that is a wrong conclusion.

Let us suppose that we are to design the script of the personal assistant, PA, that is the specialization of the Alter Ego in charge to give a warning to the cybernaut when the borrowed book must be returned.

The first kind of behaviour of the PA would be just acting as the mailbox for its owner. The library system, let us call it L, sends messages to the user about returning the borrowed book, by sending it to the PA. In this case PA's responsibility is very low. Actually, there is no responsibility, and if there were any this can be compared to the responsibility of a mail box: it should be open for messages and it should warn its owner.

In the next example we change the borrowing-a-book into borrowing-a-digital-book (or video, or cd). Because the book is digital the rule of returning it within three weeks does not make sense. We assume therefore that the user has to pay a fee within three weeks.

In a diagram, very similar to Fig. 2, this situation can be defined. The only change is that a fee has to be payed instead of book to be returned. Also a similar rule can be chosen when the fee is not payed and a fine is to be payed. In the new situation there is more to be done by the PA. Upon the request of the student/cybernaut her PA asks L for the particular book. Suppose it is available (even digital books can be non present!), suppose the student is entitled to borrow a book (remember that in the previous example she could have been blocked), then it gets the book from the library. The PA then just waits for three weeks (minus one day) and sends an order to the bank B, i.e. to its bank manager, to transfer the fee to L's account. Because this behaviour is fully defined we say that paying the fee is a NEC action. Because there is no real person anymore who can just choose to violate the commitment, there seems to be no reason to take such a possibility into account.

However, the situation is more complex. It could very well be possible that the order for money transfer is not successfully carried out, for example because the account was not enough. So what then? The only thing which ultimately is left is that the PA goes to its owner and explains the difficulty. That means that in the end the owner herself has to take responsibility. The responsibility of the PA is only seemingly present: when unexpected problems occur it is the human being herself who is in charge and not the PA. The situation is very similar to the "Call the manager" rule in many organizations.

The final conclusion therefore is that although objects cannot take real responsibility we still have to seriously take into account that they do not do what has been promised.

References

[Burg 95a] J.F.M. Burg and R.P. van de Riet. COLOR-X Event Model: Integrated Specification of the Dynamics of Individual Objects.The Proceedings of the Fourteenth International Object-Oriented and Entity-Relationship Conference (OO-ER'95), Lecture Notes in Computer Science (1021), pages 146-157, Gold Coast, Australia, 1995. Springer-Verlag.

[Burg 95b] J.F.M. Burg and R.P. van de Riet. COLOR-X: Linguistically-based Event Modeling: A General Approach to Dynamic Modeling.The Proceedings of the Seventh International Conference on

Advanced Information System Engineering (CAiSE'95), Lecture Notes in Computer Science (932), pages 26-39, Jyvaskyla, Finland, 1995. Springer-Verlag.

[Burg 95c] J.F.M. Burg and R.P. van de Riet. COLOR-X: Object Modeling profits from Linguistics. Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing (KB&KS'95), pages 204-214, Enschede, The Netherlands, 1995. IOS Press, Amsterdam.

[Dehne 94] F.Dehne and R.P. van de Riet. A Guided Tour through Mokum 2.0. Technical Report IR-368, Vrije Universiteit, Amsterdam, 1994.

[Dik 89] S.C. Dik. The Theory of Functional Grammar. Part I: The Structure of the Clause.Floris Publications, Dordrecht, 1989.

[Erickson 96] T.Erickson. The World-Wide Web as Social Hypertext, Comm. of the ACM,39(1):15-17, January 1996.

[Miller 95] G.A. Miller. WordNet: A Lexical Database for English. Comm. of the ACM, 38(11), Nov. 1995.

[Olivier 95] M.S. Olivier. Self-protecting Objects in a Secure Federated Database. Proceedings of the IFIP WG 11.3 9th working conference on Database Security, pages 23-38, 1995.

[OMG 91] OMG. The Common Object Request Broker: Architecture and Specification. Technical Report 91.12.1, OMG, December 1991.

[vdRiet 94] R.P. van de Riet and J. Beukering. The Integration of Security and Integrity Constraints in Mokum. Database Security, VIII, Status and Prospects, pages 223-246. IFIP, North-Holland, 1994.

[vdRiet 95] R.P. van de Riet, A. Junk, and E. Gudes. Security in Cyberspace: A Knowledge-Base Approach. Technical Report IR-398, Vrije Universiteit, Amsterdam, 1995.

[vdRiet 96] R.P. van de Riet, E. Gudes, An Object-Oriented Database Architecture for providing High-level Security in Cyberspace, to be presented at the IFIP WG11.3 Workshop on Security and Databases, Como, 1996.

[Varadharajan 95] V.Varadharajan. Distributed Object System Security.Information Security - the next Decade, pages 305-321. Chapman & Hall, 1995.

A Smart Curriculum for Middle-School Science Instruction: A Web-Based Curriculum Integrating Assessment and Instruction

Cognition and Technology Group at Vanderbilt
Vanderbilt University
United States of America
vyenj@ctrvax.vanderbilt.edu

Abstract: This paper discusses a model of integrated instruction and assessment called SMART (Special Multimedia Arenas for Refining Thinking). SMART involves interactive use of the internet and multimedia software. The internet serves 3 important functions: It acts as a formative assessment tool by providing individualized feedback to students, creates a learning community by displaying data submitted by participating classrooms; and promotes discussion and reflection on important concepts. We discuss design features of the SMART model, including the internet tools, and describe an example from middle-school science.

A primary focus of current research on alternative assessment involves the investigation of performance assessments for summative purposes, in particular for school- and system-level accountability [Baker & O'Neill 1994]. In this context, researchers have largely been concerned with the psychometric properties of performance assessments, notably, with issues that have arisen with respect to scoring, reliability, and validity.

While these issues are important, this paper concerns a different, often overlooked, function of assessment--to inform ongoing instruction and learning. There are several reasons for our focus on formative assessment[1]. One derives from our goal to transform classrooms into exciting learning communities that encourage students to achieve high levels of learning. We assume that teachers will make better instructional decisions by engaging in formative assessment, leading to improved student learning. We also assume that students will learn more and learn more deeply if they routinely engage in reflection and revision [Brown 1987].

In addition, formative assessment is fundamental to the new standards for instruction recommended by groups such as the [National Research Council 1996] and the [National Council of Teachers of Mathematics 1989]. These standards, based on social constructivist principles of learning, emphasize the importance of teaching in ways that promote deep understanding by students. Learning is no longer viewed as the accretion of new information. Instead, it is viewed as transformational process wherein concepts slowly evolve. In this milieu, teachers are being asked to adopt a more "cognitive" stance to teaching, for example, to be aware of the preconceptions that their students bring to new learning situations, to teach in ways that make students' thinking "visible" to other students and to help students reflect on and reconcile their conceptions with those of others. Formative assessment is an important part of this new repetoire of teaching behaviors.

In this paper we discuss a model of integrated instruction and assessment that we call SMART (Special Multimedia Arenas for Refining Thinking). SMART involves interactive use of the internet and multimedia software for assessment and instruction. In SMART the internet functions as both a teaching tool and an assessment tool. We first describe design features of the SMART model, and then discuss an example from the area of science. In this context we elaborate on details of the tools we developed for the internet.

---

[1] We use the term "formative assessment" to refer to reflective practices by both teachers and students. Formative assessment by students is synonymous with self-assessment activities where students reflect on their conceptions. Formative assessment by teachers involves ongoing monitoring of students' knowledge and skills for purposes of instructional decision-making.

EVOLUTION OF THE SMART MODEL

The current SMART model derives from close to 10 years of research with students and teachers on ways to motivate and assess exceptional learning [Barron, Vye, Zech, Schwartz, Bransford, Goldman, Pellegrino, Morris, Garrison, & Kantor 1995] [Cognition and Technology Group at Vanderbilt 1994]. Our initial work focused on mathematics, and was concerned with an approach to instruction that we call "anchored instruction." In anchored instruction, teaching and learning are focused around complex problems or "anchors." The anchors are stories on videodisc (or CD-ROM) that each end with a challenge to solve. All of the data needed to solve the challenges are contained in the stories. The problems (a) are complex and require extended effort to solve (at a minimum, in the range of 3-5 hours for most middle school students); (b) are relatively ill-defined and require significant formulation prior to solving; and (c) have multiple viable solutions. The anchors are designed to engage students in authentic problem solving activities that highlight the relevance of mathematics to the world outside the classroom.

Our initial research on anchored instruction was conducted using "The Adventures of Jasper Woodbury", a series of video anchors developed by members of our Cognition and Technology Group at Vanderbilt. Findings indicated that working on multiple Jasper anchors over the course of a school year resulted in significant improvements in fifth and sixth grade students' problem formuation and problem solving skills. In addition, students showed positive changes in their attitudes towards mathematics [Pellegrino, Hickey, Heath, Rewey, Vye, & Cognition and Technology Group at Vanderbilt 1991]. Nonetheless, reports from teachers and students were unanimous in their strong dislike for our assessments (these assessments were conducted as part of our research and consisted of traditional paper and pencil mathematics story problems).

In thinking about how to re-design our assessments, we focused on assessment as it occurs outside of school settings. This was a valuable thought experiment in that it pointed to some important differences between assessment in and outside of school. First, in contrast to assessment in schools, assessment in professional contexts is usually external, and the products that are assessed are designed to contribute in some way to the profession. For example, when we write a paper or prepare a proposal for a presentation, our work is examined by expert individuals who are external to our department. Further, evaluation is not the sole purpose for generating products. Hopefully, papers and presentations contribute to knowledge, research and development in the field. In designing SMART we have tried to emmulate these features. Students' learning is directed toward culminating challenges that are evaluated by experts and have tangible consequences. For example, in some of our early work, students from different classes participated in live satellite programs in which they responded in real time to challenges related to Jasper [Kantor, Moore, Bransford, & Cognition and Technology Group at Vanderbilt 1992]. In recent work, the culminating challenges relate to project activities that follow Jasper. In one of the Jasper anchors students learn to design blueprints and a scale model of a playground and playground equipment, and in the project that follows, students are challenged to design blueprints and a scale model of a playhouse for kindergarten-aged children. Students present their designs to expert builders, and designs that meet prespecified evaluation criteria are entered into a random drawing. Designs selected during the drawing are built and donated to local kindergarten classes.[2]

Another way in which assessment in school and professional settings differs relates to opportunities that are available for improving one's work. In professional settings, there is a comittment to creating the very best product that is possible. We solicit input from people both internal and external to our organization, and we pay careful attention to performance standards set by experts in the field. We rely on this information as we draft and re-draft our work; reflection is a critical part of the process. Regrettably, in most classrooms, opportunities for feedback, reflection, and revision are almost non-existent. When students do receive feedback, it is usually in the form of a grade--rather than something that could help them enhance their understanding--and opportunities to improve their work are rare.

To promote self-assessment and reflection in SMART classrooms, instruction is explicitly organized around cycles of work and revision, and we have designed technology-based tools to that provide feedback to students and help them improve their work. Our research indicates that students who use these tools learn significantly more than students who go through the same instructional sequence for the same

amount of time, but who do not use the tools [Barron et al. 1995]. Initially, our tools consisted of videodisc programs and stand-alone computer applications. More recently, we have used the internet. In the sections below, we describe a just-completed experiment using our SMART WWWeb.

---

[2] Space constraints preclude a comprehensive review of our research on SMART. We refer interested readers to [Barron, et al. 1995] and [Cognition and Technology Group at Vanderbilt 1994] for more information.

THE WEB AS A TOOL FOR TEACHING AND ASSESSMENT

As mentioned, in SMART students iterate through cycles of problem solving and revision. Students access our internet site, SMART WWWeb, during the revision phase. Essentially, SMART WWWeb serves 3 functions: First, it provides individualized feedback to students. In this way, the Web serves as a formative evaluation tool. The feedback suggests aspects of students' work that are in need of revision, and classroom resources that students can use to help them revise. The feedback does not tell students the 'right answer.' Instead, it sets a course for independent inquiry by the student. The Web feedback is generated from data that individual students enter. Essentially, data that is submitted by students in the brower is collected in a database on our server. Responses in the database are subsequently tagged with feedback that is sent back to the browser for students to print out.

The second function of SMART WWWeb is to collect, organize and display the data collected from the distributed classrooms. Data displays are automatically up-dated as new data are submitted to the database by students. We call this section, SMART Lab. The data in SMART Lab consist of students' answers to problems and explanations for their answers. Each class' data are displayed separately from the distributed classroom's data. This feature enables the teacher and her/his class to discuss different solution strategies, and in the process, address important concepts and misconceptions. These discussions provide a rich source of information for the teacher on how her/his students are thinking about a problem, and are designed to stimulate student reflection as well.

The third section of SMART WWWeb is Kids Online. Kids Online consists of explanations by student-actors. The explanations are text-based with audio narration. Still pictures of the presenters are also available. The explanations are errorful by design. Students are asked to critically evaluate the explanations,

and provide feedback to the student-actor. By including errors we are able to seed thinking and discussion on concepts that are frequently misconceived by students. At the same time, students learn important critical evaluation skills.

AN EXAMPLE FROM SCIENCE

Our current work on SMART is focused around a video anchor on CD-ROM entitled, 'Stones River Mystery' (hereafter SRM). SRM is an episode in the series "Scientists in Action" developed by Bob Sherwood and his colleagues at Vanderbilt [Sherwood, Petrosino, Lin, Lamon, & Cognition and Technology Group at Vanderbilt 1995]. SRM tells the story of a group of high school students who, in collaboration with a biologist and hydrologist, are monitoring the water in Stones River. The video shows the team visiting the river and conducting various water quality tests. (In our work we focus on benthic macroinvertebrate sampling and dissolved oxygen testing.) Students in the classroom are asked to assess the water quality at a second site on the river. They are challenged to select tools that they can use to sample macroinvertebrates and test dissolved oxygen, to conduct these tests, and to interpret the data relative to previous data from the same site.

When students begin working on macroinvertebrates, they are given a catalog of sampling tools/instruments. Many of these are "bogus" and collect the wrong kind of sample; others are 'legitimate' and will gather a representative sample of macroinvertebrates. The catalog items are specially designed to include contasting cases that help students discover the need to know certain kinds of information. For example, to use macroinvertebrates as indicators of water quality, one needs to collect a sample that represents the river's biodiversity, and as such, all types present need to be sampled. The following is the actual description of one of the items in the catalog, the TetraBen Laser Counter:

"Knowing the number of macroinvertebrates in your water is an important way to determine the health of your river. Collecting and counting these organisms can be a slow, tedious process. Modern science has revolutionized this process. The TetraBen Laser Counter lets you count macroinvertebrates without getting your hands wet! Simply scan the laser beam slowly over the water. The laser beam automatically counts the macroinvertebrates, and shows the total number on a built-in screen. The laser is completely waterproof and won't harm anything, living or non-living (and that includes macroinvertebrates and humans!) Simple, safe, and completely accurate!"

This is an example of an item that would collect the wrong kind of information; it counts the macroinvertebrates and ignores information about the types of macroinvertebrates in the sample.

Students are asked to choose and justify their choice of tool. To help them make their choices, they are provided with resources, some of which are on-line, that they can use to find out about river ecosystems, macroinvertebrates, and water quality monitoring. Once students have made an initial set of choices, they use SMART WWWeb. They enter their catalog choices (yes or no), and select justifications for their choices (why or why not choose that catalog tool). Figure 1. shows a portion of the internet "order form."

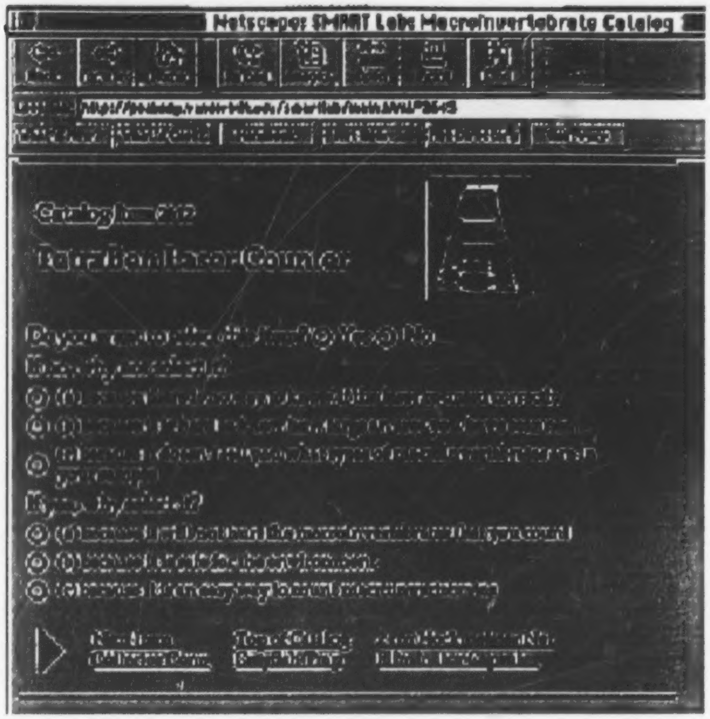
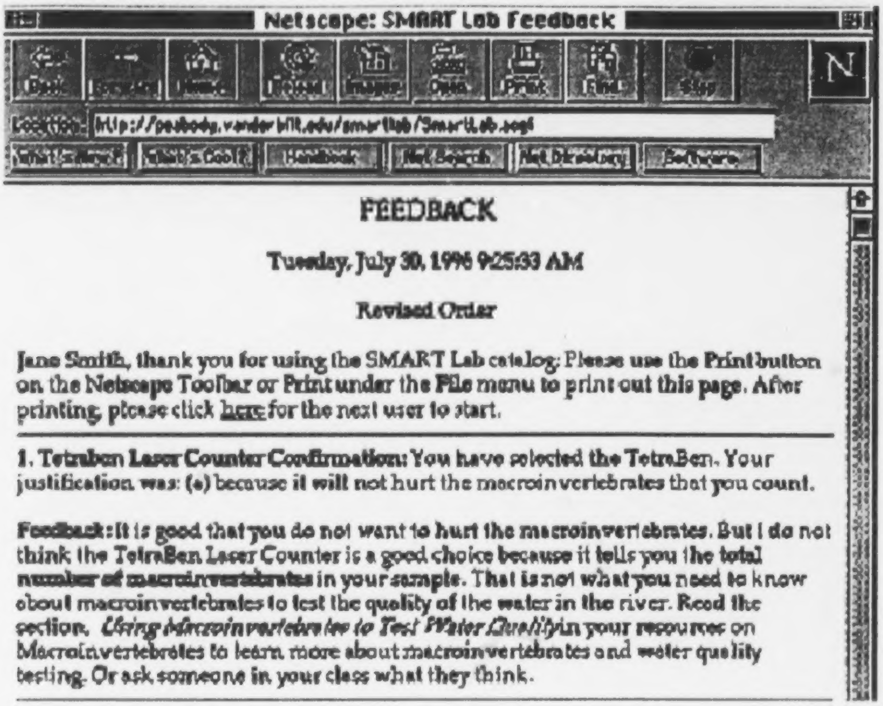


Figure 1: Excerpt from SMART WWWeb macroinvertebrate catalog

An important feature of the catalog and internet order form is that they are designed to reveal common misconceptions about ecosystems and pollution. For example, our pilot research on science showed that many students think that bacteria is harmful and pollutes rivers. We have tried to include catalog items and foils that expose particular misconceptions. In our macroinvertebrate catalog, we have a 'Super Collector Cone' that promises to collect macroinvertebrates and bacteria, and if students decide to order the Cone, one of the justifications that they can use is that 'bacteria pollute the water so it is important to catch them.'

Once students have submitted their catalog order on-line, SMART WWWeb sends them individualized feedback. Figure 2. shows a segment of SMART WWWeb feedback.

BEST COPY AVAILABLE



FEEDBACK

Tuesday, July 30, 1996 9:25:33 AM

Revised Order

Jane Smith, thank you for using the SMART Lab catalog. Please use the **Print** button on the Netscape Toolbar or **Print** under the **File** menu to print out this page. After printing, please click here for the next user to start.

1. Tetraben Laser Counter Confirmation: You have selected the TetraBen. Your justification was: (a) because it will not hurt the macroinvertebrates that you count.

Feedback: It is good that you do not want to hurt the macroinvertebrates. But I do not think the TetraBen Laser Counter is a good choice because it tells you the total number of macroinvertebrates in your sample. That is not what you need to know about macroinvertebrates to test the quality of the water in the river. Read the section, *Using Macroinvertebrates to Test Water Quality* in your resources on Macroinvertebrates to learn more about macroinvertebrates and water quality testing. Or ask someone in your class what they think.

Figure 2: Excerpt of student feedback from SMART WWWeb

The feedback that students receive from SMART WWWeb highlights why the selected tool is problematic, and suggests helpful resources (sections of on-line and off-line resources, hands-on experiments, and peers). This form of feedback is similar to the feedback that we used in our work on mathematics, and that our research suggests can be an effective stimulus for guided inquiry and revision by students.

SMART Lab and Kids Online are accessed next. Teachers use these sections to engage their classes in critical discussions. In the process, teachers and students discover how class members are thinking about issues. For example, SMART Lab summarizes the catalog choices and justifications submitted by each class and displays these data with the same data aggregated across all classes. Classes can discuss how their data are similar and different from the distributed class' data. Or they can discuss whether they agree with the most popular reasons given for choosing catalog items. In Kids Online, student-actors discuss their own catalog choices. We purposefully include reasoning errors in these presentations, and target common misconceptions that students have about river science. In the course of discussing Kids Online, students confront and hopefully debunk these misconceptions.

After visiting the internet, student work to revise their catalog selections. They have opportunities to look up text-based resources that provide more in-depth information about science content relevant to various choices. For example, the text resources explain the need to break macroinvertebrates into categories, namely, pollution tolerant, somewhat pollution intolerant, or pollution intolerant. Students can use this information to understand why the TetraBen Laser Counter, which counts all macroinvertebrates but does not sort them, does not provide the kind of data they would need.

After revising their thinking, students again visit the SMART WWWeb and make new choices of catalog items and new justifications. They can then see summarized data from their class and other classes and see how the data have changed. Following the correct choice of an appropriate object (in this case the .5mm Hochmeister Kick Net), students work with a CD-ROM simulation that allows them to 'see' a sample of macroinvertebrates, calculate a water quality index, and compare their results with baseline data from previous years. Although each student gets a different sample of macroinvertebrates, each set of data shows that there is a serious absence of pollution sensitive macroinvertebratesóhence something is wrong.

The SMART Challenge continues by next having students choose items for doing an oxygen test. Again, they make their choices via the Web and see data that summarizes the choices of other classes. Also, students gain access to text-based resources (which sometimes reside in other internet sites) that help them understand the science underlying various choices. And they eventually do some experiments on their own. For example, students are encouraged to test the amount of dissolved oxygen in a tank of water prior to putting fish in it and after the fish have lived in it for at least one day. With appropriate testing instruments, data show that there is less dissolved oxygen in the water after the fish have been there. For classrooms that cannot do actual tests, simulated, computer-based tests are available. All of these activities are preliminary to project-based activities in which students conduct water quality testing at a local river and present and publish their findings for the local Water Quality Management Department.

As noted above, our work with SMART WWWeb challenges is just beginning: The Stones River Mystery challenge is the first we have attempted. By providing students and teachers with frequent opportunities for formative assessment and revision, we believe that we can better help them reach the goals of the National Standards in areas such as science and mathematics. Our plans are to create additional internet-based challenges that focus on 'big ideas' in areas such as mathematics, social studies, and literature.

REFERENCES

[Baker & O'Neill 1994]. Baker, E. L., & O'Neill, H. F. (1994) *Technology assessment in education and training*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[Barron, Vye, Zech, Schwartz, Bransford, Goldman, Pellegrino, Morris, Garrison, & Kantor 1995]. Barron, B., Vye, N.J., Zech, L., Schwartz, D., Bransford, J.D., Goldman, S.R., Pellegrino, J., Morris, J., Garrison, S., & Kantor, R. (1995). Creating contexts for community-based problems solving: The Jasper Challenge Series. In C. Hedley, P. Antonacci, & M. Rabinowitz (Eds.), *Thinking and literacy: The mind at work* (pp. 47-71). Hillsdale, NJ: Lawrence Erlbaum Associates.

[Brown 1987]. Brown, A. L. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In F. Weinert, R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, NJ: Lawrence Erlbaum Associates.

[Cognition and Technology Group at Vanderbilt 1994]. Cognition and Technology Group at Vanderbilt (1994). From visual word problems to learning communitites: Changing conceptions of cognitive research. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 157-200). Cambridge, MA: MIT Press/Bradford Books.

[Kantor, Moore, Bransford, & Cognition and Technology Group at Vanderbilt 1992]. Kantor, R. J., Moore, A. L., Bransford, J. D., & the Cognition and Technology Group at Vanderbilt. (1993, April). *Extending the impact of classroom-based technology: The satellite challenge series*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

[National Council of Teachers of Mathematics 1989]. National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

[National Research Council 1996]. National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

[Pellegrino, Hickey, Heath, Rewey, Vye, & Cognition and Technology Group at Vanderbilt 1991] Pellegrino, J., Hickey, D., Heath, A., Rewey, K., Vye, N. J., & Cognition and Technology Group at Vanderbilt (1991). *Assessing the outcomes of an innovative instructional program: The 1990-1991 implementation of the 'Adventures of Jasper Woodbury'* (Tech. Rep. No. 91-1). Nashville, TN: Vanderbilt University.

[Sherwood, Petrosino, Lin, Lamon, & Cognition and Technology Group at Vanderbilt 1995]. Sherwood, R. D., Petrosino, A. J., Lin, X., Lamon, M., & the Cognition and Technology Group at Vanderbilt (1995). Problem-based macro contexts in science instruction: theoretical basis, design issues, and the development of applications. In D. Lavoie (Ed.), *Towards a cognitive-science perspective for scientific problem solving* (191-214). National Association for Research in Science Teaching: Manhattan, KS.

Acknowledgements

The preparation of this chapter was supported by a grant from the National Science Foundation (NSF MDR-9252908). Members of the Cognition and Technology Group at Vanderbilt who contributed to the work described in the paper are: Brigid Barron, Helen Bateman, Kadira Belynne, John Bransford, Joan Davis, Michael Gaines, Susan Goldman, Susan Hickman, Michael Jacobson, Taylor Martin, Cynthia Mayfield-Stewart, Jim Pellegrino, Dan Schwartz, Carolyn Stalcup, Nancy Vye, and Linda Zech.

A Course in Internet / Multimedia for Non-Computer Science Majors

Jerry Waxman
Department of Computer Science
Queens College
Flushing, NY 11367
waxman@aleph.cs.qc.edu

Abstract: The second course for non-computer science majors is presented . The course focuses on the acquisition and presentation of data and information. For data acquisition, the course will give the students high level mastery of the Internet. For the presentation of information the students will learn multimedia authoring such as Toolbook and Director, but the primary focus will be on Web publishing. The course will be structured around (1) a lecture which stresses high level mastery of the software tools, (2) exploratory, hands-on supervised lab sessions in a specially constructed lab, (3) assignments which require students to continue, on their own, the exploration begun in the lab, and (4) a support network of graduate and undergraduate student aides.

Background

At Queens College, we offer a novel first course in computing for non-computer science majors. This course, CS10, was developed with the support of FIPSE (U.S. Department of Education ,1990-1993) and the ILI program of NSF (1989), and focuses on teaching students who are mostly non-technical, to become high-level "power users." The course provides two hours of lecture per week and an additional two hours of an intensive supervised lab/recitation where the students work on "lab experiments" to explore the features of various software packages. The NSF funds were used to build the lab and the FIPSE funds supported the curriculum development.

This course differs significantly from other "software tools" courses in that students not only learn how to use software "tools" such as wordprocessors, spreadsheets and databases, but they learn how to **program** in those environments as well, using the macro and programming tools provided by those packages. The course stresses the idea that all software is "model-based", and by discerning the model underlying the software, different packages can be more easily mastered. Students are not only taught the use of specific programs; but how to master new software by themselves. Computer science concepts such as algorithms and programs are presented in a useful context. They are better understood and assimilated than if taught abstractly or in contexts that are less useful for the non-major, such as Basic or Pascal.

The course has been a resounding success with the **more than four thousand students** having completed the course. Before its redesign, the course for non-majors had an attrition rate of 30 to 40 percent, mostly women and minority students. Now the attrition rate is negligible. Despite the demanding workload, the course is always over-subscribed, increasing each semester. As **a direct result of their experience in the course**, students seek out other computer and mathematical courses. Students comment on the boost the course provided them in seeking employment. CS10 has been designated by the college curriculum committee as fulfilling a college-wide requirement for a course in "quantitative reasoning." Many departments have required the course as part of their major.

CS20

Due to the success of CS10, there has been considerable interest in a **second high-level course in computing for non-majors.**. We are developing a course which focuses on the **acquisition and**

presentation of data and information. For data acquisition, students will gain high level mastery of the Internet. For presentation, they will learn multimedia authoring with a primary focus on Web publishing. This course is for those who completed CS10, or can demonstrate competence. While in the past, only science and engineering students needed high level computing skills, fluency with computational environments will be a decisive factor in academic and commercial success.

CS20 will provide students with new techniques and paradigms, allowing them to bring new technological know-how to bear in solving problems **in their own** areas. The focus of CS20 will be on techniques **for acquiring, organizing and presenting that data**. CS20 will provide the general skills required to locate data in and transform it into information. In concentrating this material in CS20, as opposed to within the context of the subject matter, many more resources, techniques, and systems can be covered. Subject matter experts can present their material at a much higher level because the students would have a solid grasp of the "information universe" and the purely technical "how to." CS20 will be structured with the following three parts:

1. **Sources and techniques for global data acquisition** with a primary focus on the structure and use of the Internet.
2. Mastering the concepts, software and methods for use in **presenting information**, particularly the use of **multimedia authoring systems**.
3. Mastering the concepts, software and methods for use in **publishing on the World Wide Web**.

In learning to navigate Internet, students gain access to a vast reservoir of data and information that will transform the kind of work they are able to do. Psychology students can access the "alt." news groups in Netnews, pose questions, and even set up surveys with an international call for responses.

While some systems such as Netscape and Mosaic are available to help navigate the net, "power surfers" will still need to know Telnet, Archie, ftp, WAIS, Gopher, and finger utilities. Thus the first component of the course, Internet concepts and tools, will focus on the following topics:

- What is the Internet and what is it not.
- Resources available on the Internet for teaching and research in various fields.
- Navigating the Internet via Gopher, Archie, Veronica
- WAIS, WWW with Netscape and HTML
- Email (with MIME), FTP, Telnet and Listservers
- Netnews, netiquette and ethics
- Student projects

Instruction goes beyond the fundamentals and aim for the level of "advanced user/ mini-wizard", including the knowledge required to setup and maintain their own web sites and lists.

A Stress on Principles

Since specific software packages evolve and / or are exchanged for newer more powerful ones, **principles** occupy a central place in the course. Practical knowledge is presented within the framework of an understanding of computer networks and communications systems principles. Issues of privacy and the security of information will be covered in depth, as will ethical questions arising in the context of computation and communication. This component will take up one third of the course.

In terms of data presentation, the principles underlying hypertext and hypermedia will be taught through use of multimedia authoring systems (on and off the Web) including **Toolbook** by Asymetrix Corp. and Macromedia's Director software. These systems use different metaphors for the organization and presentation of hypermedia systems. The students will be introduced to multiple paradigms for multimedia construction.

In terms of practical skills, students will learn to create sophisticated interactive presentations which integrate text, sound and video. Specifically they will learn:

- how to construct simple "books" with hot keys and buttons
- how to create multi-page branching books
- how to create motion and animation
- how to design and create hypertext systems
- how to design and create interactive multi-media
- how to **program** "sophisticated" multimedia applications within Toolbook and Director Toolbook contains an extensive and powerful programming environment, and Director is built around the Lingo programming language. Programming projects, in their area of concentration, will reinforce the programming and analysis skills acquired in CS10. This will be one third of the course.

The final 20 hours of the course will be an introduction to publishing on the World Wide Web. The Web provides the ability to access hypertextually organized documents with links to other documents, images, sound and video clips. With the vast storage potential of the 5 million "hosts" and the effective distribution of computation over the network, it will be possible to write applications of extraordinary beauty and utility. More and more companies are hooking up to the Web and creating "home pages." The demand for people to design applications in this context greatly outstrips the available talent. Students will learn to design and implement fairly sophisticated Web applications. The stress will be on principles and their illustration in the context of contemporary tools. In the second section of the course, students will learn the principles of hypertext within the simpler context of the CD-ROM environment. The Web adds a whole new set of conceptualizations and perspectives. Students need to learn to think in network terms. In this context, multiple threads, timing constraints, data transmission volumes and network bandwidth become important considerations.

How the course will be structured

CS20 is being structured similar to CS10. Twice weekly lectures will be complemented by twice weekly supervised labs. Lecture and laboratory manuals are being prepared and will be revised periodically based on an ongoing evaluation of their effectiveness. The recitation will be devoted to working out **numerous examples** and to giving the students the much needed **supervised hands on experience**. Difficulties can be dealt with by the recitation instructor. Interesting problems and their solutions may be shared.

What will be the impact of the course.

CS20 will significantly improve the quality of computer science instruction for non-majors. Within the next year, enrollment could be 200 students per semester. The limiting factor is the lack of adequate lab space. We have applied to various agencies to help fund facilities devoted to just CS20.

There has been a growing recognition of the need for a **new type of computing minor**. Currently, the computer science minor focuses on the needs of students majoring in the sciences, engineering or mathematics. There is a need for high level computing skills in many other disciplines. CS10 and CS20 courses could form the core of such a minor. Many students currently in this course are Education or Communications majors. Empowering large numbers of people with these skills will have a profound impact on both the educational and business communities.

The need for such a course is not unique to Queens College. We hope and expect that the techniques and course materials developed at Queens could be used by institutions all across the country.

EPIC : Building a Structured Learning Environment

Martin D. Westhead
Edinburgh Parallel Computing Centre
University of Edinburgh
U.K.
martinwe@epcc.ed.ac.uk

Abstract

Organisation and structure in the space around us are used to reduce the amount of memory and decision making that must be applied to carry out tasks in everyday life. This paper outlines work in progress at the University of Edinburgh on the construction of a small web based interactive learning environment (EPIC) developed for the teaching of high performance computing. The paper begins by outlining work done in Cognitive Science on how people make use of structure in physical environments, within this context the EPIC system is presented as an illustration of how the principles can be applied to aid the learning process.

1. Introduction

Recent work in Cognitive Psychology [Kirsh95] demonstrates how important the structuring of environments can be in carrying out tasks. The work draws on studies of everyday observations in kitchens, supermarkets and workshop and shows how, by structuring their environment, people simplify the reasoning, memory and perception which are required.

Learning is an activity which is very demanding in terms of cognitive load. It is proposed that the structuring of a learning environment can alleviate unnecessary aspects of that load, and allow the learner to focus on the material being presented. By way of illustration a case study is presented of work carried out at the Edinburgh Parallel Computing Centre (EPCC), on a Web-based learning environment. The system, called EPIC (EPcc Interactive Courseware), extends the hypertext environment of the Web to provide an simple, consistent interface to exercises on a number of courses.

2. Structuring Environments

In his paper on intelligent use of space David Kirsh [Kirsh95], develops the beginnings of a principle classification of the way in which people make use of features of the space around them to simplify tasks. We are embodied agents and the nature of our interaction with the world about us is inherently limited. We face in one direction at a time and we can only immediately manipulate objects within our reach. The space about us also conforms to physical and geometric laws. We use these constraints to our advantage and in this section we look at some of the ways in which this is done. Kirsh argues that whether we are aware of it or not, we are constantly organising and reorganising our environments to enhance our performance.

Some of the techniques we use are obvious, but others are learned, and developed over time. Experts in particular excel at organising their tools and materials so as to minimise the number of decisions which they have to have to make in performing a task. For example a short order cook given an order for say, mushroom omelet and toast, will begin by laying out their workspace. Putting the eggs by a bowl with a fork, laying out the mushrooms and bread by a chopping board. Thus in the cooking stage, activity can progress from one item to another at high speed without worrying about the higher level planning involved. Kirsh reports that in cases where several orders are being prepared at the same time the cooks will cluster materials for orders together and leave utensils by ingredients to mark their purpose.

Even in everyday tasks people use spatial ordering to cut down the performance of the task to a series of unreflective actions, triggered by environmental cues. Making a cup of tea in your own kitchen will proceed automatically in a skill based mode of activity. You don't need to think about where the cups, tea kettle are, or how to arrange them. Simple problems are solved with "problem solving packets" determined by rules. Finding that there are no clean mugs you will "automatically" go and wash one. It is only when the task gets out of control that conscious analytical thought needs to be applied.

In order to increase the amount of time a person spends in skill or rule based modes, they can either broaden their range of skills and rules, or they can structure their environment to limit the uncertainty. Such structures can involve physically limiting options or providing perceptual cues to act as reminders. Greying out items on a pull down menu which are not available for selection is one example, another might be placing a library book by the door, as a reminder to return it when you go out. If such cues are successful they capture our attention at critical times and either alert us to opportunities or reduce the likelihood of "double-capture slip", this occurs when we are distracted from a course of action and habit takes over. e.g. "I intended to stop on the way to work to buy shoes, but 'woke up' to find I had driven right past."[Reason82,90]

We will notice that an action is available in the current situation if the environment *affords* that action. An affordance in a particular situation can be described as the way a situation lends itself to being used [Gibson77]. Thus a button affords pushing, a bicycle affords riding, an empty glass affords filling. The affordance offered by a knife and chopping board next to unchopped mushrooms by our fast cook chief is clear. It is important to see that the affordances of an object are features of the agent and the environment, if you cannot ride a bicycle, or if it is chained up, it will not afford riding. In structuring our environment we both hide and highlight affordances to make to change the likelihood of them being chosen.

The structure of an environment is not always the result of an individual's organisation, structure may also be imposed. A kitchen on its own has a great deal of physical structure associated with its use, large sinks, cookers, work surfaces, cupboards etc. Another good example of imposed structure is an assembly line, where efficiency is increased by limiting the tasks an individual performs, and ordering the tasks in a linear fashion.

In summary we exploit the constraints we face as embodied agents interacting with the world about us to simplify the execution of tasks. The next section looks at the EPIC package to illustrate how we can begin to use these principles in a learning environment.

3. EPIC

EPIC is a general purpose interactive learning package developed at EPCC. As a leading centre for High Performance Computing (HPC) in Europe, EPCC has a number of different education and training requirements. Within the centre, the education and training group is responsible for the training of staff, users of our facilities, visiting researchers and industrialists, summer students, as well as a wider remit to bring HPC to Higher Education Institutes throughout Scotland and Northern England. With such diverse requirements we are involved not only in classroom based courses, but also in stand-alone course material, and distance learning.

3.1 The EPIC system

Many of the courses taught at the centre focus on learning programming languages. This involves the users iterating through a cycle of editing their code, compiling it and running it, returning to the editing stage when things fail. Since this edit-compile-run cycle was at the heart of many of the courses we taught, it seemed like an obvious place to start providing interaction. The World Wide Web by its ubiquity and high quality, free, GUI supporting software was an obvious choice of platform, particularly with distance learning being one of our goals.

Unfortunately the Web was not really designed for interactive courseware. One of the obvious drawbacks being that service is provided on a document, by document basis and no record of a user is maintained. This `lack of state' makes it difficult to maintain an interaction. There are a number of mechanisms that can be used to achieve the desired result. For example state can be maintained in forms which interact with smart CGI scripts, however we felt that a forms based interface was too limiting. Instead we made use of the external viewer mechanism in most browsers, which causes the browser to pass a page on to an application, if it is of a defined MIME type. We defined a new MIME type for EPIC files will cause a browser to execute a script on the client. Using this mechanism we can create exercise pages with buttons that launch processes on the client machine.

The buttons on an typical exercise page enable the user to:

- launch an editor of their choice, containing the code skeleton associated with the current exercise,
- fire up an interactive shell window which automatically compiles the appropriate file,
- run the executable produced in the same window,
- view the results using a visualisation package,
- clear up all the windows,
- provide configuration facilities, such as changing editor preferences, or number of processors in a run,
- and finally amongst the security measures, is a monitor window which reports all the actions EPIC takes. The actions are strictly limited by the script, but this ensures that the scripts are carrying out the action the user expects.

The EPIC exercises are embedded within a complete set of hypertext course-notes. Instructions to users on commands or techniques in the course can be linked back to the place in the notes where they were first encountered. Links can also be made to other relevant online material such as language specifications, or FAQs about the course.

3.2 EPIC as a Structured Learning Environment

Although it is cute to be able to run the compiler from a button on the exercise page, it could be argued that the EPIC package is in fact doing nothing new. Students on a course could do all the things that EPIC allows them to do just by typing in the right command line. However by the same argument the Web itself does nothing new, long before the Web we had anonymous ftp sites. You could go to a site an download documents, pictures, audio, video etc. However the Web makes the whole process much less painful. It provides structure to cyberspace. It relieves the user of the burden of logging on, finding the remote file, copying it somewhere sensible in his or her filespace uncompressing it, unarchiving it, displaying it, and then once read potentially deleting it. The Web presents links which afford clicking, in a context which provides cues to guide browsing, all within an environment with a consistent `look and feel'. Although it is true to say that the Web does nothing that couldn't be done before, it is also true that the cognitive burden which is relieved by this structure, qualitatively changes the experience of retrieving files from the Internet (bandwidth notwithstanding). EPIC extends the structure made possible by the Web technology to apply it to the edit-compile-run cycle of the programming courses. In designing the system, we looked at the various tasks that a user would have to perform, and looked at how an interface could be designed to hide unnecessary details and provide affordances for these tasks. In our case the principle subtasks that we were concerned with were editing, compilation and execution. Before EPIC was introduced courses required users to copy and manage various skeleton code and configuration files that were used though the course and correctly invoke the compiler, execution mechanism, and visualisation packages. This involved remembering, or frequently looking up:

- pathnames,
- filenames,
- compiler commands,
- compiler flags,
- execution commands and
- execution flags.

Furthermore many of these details can change subtly between stages of an exercise. Although this may not sound very significant it does present a burden that someone learning a new programming language could well do without. All these details are hidden in EPIC by a button push allowing the users to focus on the exercise itself.

EPIC also seems to reduce the likelihood of "double-capture slips". It is frequently the case that you would see a user in a training course come out of an editor, or finish a compilation, and forget what they intended to do next. Habitual response takes over in these cases. In UNIX users this is often to type 'ls' and list the files in the current directory. This behaviour is quite common. In a sense the user is searching for environmental cues that will trigger them to remember the next step. In EPIC the buttons are there on the exercise page, *affording* clicking. The interface is all contained on the Web browser in a consistent style. What is more the buttons are arranged in the expected order of use "edit", "compile", "run", "view","clear-up". Since we naturally read from left to right (most course participants are from the West) the user will encounter the buttons they need in the order they will use them. Admittedly it would be better if buttons were shaded until they could be used, "run" was grayed out until "compile" had been used. Unfortunately the stateless nature of the Web interface, and the tenuous nature of the communication with the client script make this difficult to achieve.

Finally EPIC provides a consistent `look and feel' to the EPCC courses. Some of our users are unfamiliar with the Web, the interface is sufficiently simple that they quickly get the hang of using it. Since the same interface is used on all the exercises, across different courses the user need only learn how to use it once.

3.3 Use of EPIC

EPIC has been in use for over a year now. We currently have EPIC courses in MPI (Message Passing Interface) and HPF (High Performance Fortran) and the system is being routinely introduced for our classroom courses on these topics. A great deal of interest has been shown by a number of people in the software, which is now publicly available, and courses have been run using EPIC at Glasgow and Aston Universities.

EPIC has been used within the centre on around 30 users, in roughly half and half stand-alone, and classroom based courses. This represents in the region of 200 hours of use. Feedback from the users has been obtained in the form of questionnaires, and this information has been used to guide the development of the system.

On the whole feedback has been very positive. When asked what he most liked about the course, one user wrote:

"The possibility, with EPIC, to concentrate on the main parts of the development of the programs."
Other comments include:

"...one can concentrate on really learning MPI, without wasting time on compiler options."

"I was unable to attend an MPI course given by EPCC staff and had an urgent need to learn the basic MPI routines. This course enabled me to do so quickly and at my own pace."

On general course questionnaires EPIC was frequently listed as a positive feature.

Of course there were criticisms too. The response speed of the system was too slow for some. We also had two experienced programmers who found the exercise styles and interface too restrictive. Other training groups have commented that they feel that some of the details which EPIC hides it would be useful to learn, compiler flags for example.

We feel quite strongly that for our courses, these details are superfluous to what is being taught. There are many language compilers, each with different execution commands, and different flags. So in learning these details a user would only really benefit if they were using exactly the same system as EPCC. Even then these are hard to remember and are easily looked up.

In the past users unfamiliar with UNIX had to learn how to get around a new operating system in order take a course. EPIC abstracts some or all of these details away so that the course is really about the material it sets out to teach, not the vagaries of the operating system or compiler that is being used.

3.4 Future plans

The EPIC project has been about developing a working tool that we can use in the Centre. The development has involved extensive field testing and user feedback to ensure that we were building something useful. Despite the positive feedback so far the current implementation is only a small step towards what could be accomplished in providing a learning environment. However this is work in progress and we have a number of plans to build on this initial work.

The EPIC system has just reached the end of its first phase of development (at time of writing), the software has been made publicly available, and we are encouraging people to download the client and try the courses. However the project is ongoing, and there are a number of future directions in which we hope to take the system. The most significant developments will be towards improving the system's potential for distance learning. We envisage at least three possibilities:

1. Currently it is necessary for distance users to have their own copies of the compilers and other software used. This limits the distance audience. A solution would be to provide another interface which appears almost identical to the user but which causes programs to be compiled and run at EPCC. There are of course security issues associated with this, but these can be solved.
2. Interaction in the current implementation is simply interaction with a machine. It is important in the creation of a rich learning environment to extend this to facilitate interaction with other people. There are two directions we want to take this in.
    - Communication with an on-line tutor --- A user would hit a button to connect with a tutor, this would transfer log information to the tutor automatically setting up a mirror of the students page and code, and creating a chat window, or frame within the browser to allow the students to communicate directly with an expert advisor.
    - Communication between students --- We hope to set up bulletin board and mailing list to build on the experiences of people who have taken the course. This will provide another feedback mechanism to help us tune the course, and encourage the students to help each other in the learning process.
3. Taking the communication a stage further we are also looking at the possibility of using the MBone to transmit interactive courses around Scotland. This would give us video conferencing facilities, which would complement the EPIC tutoring facilities. In a sense the EPIC system would allow the tutor the electronically look over the shoulder of a student at a remote site.

The mechanism that the system uses is very general, and could be used to activate a number of other packages and applications on the user's machine. We are already developing two new courses that will use EPIC to drive applications in this way. Other include the possibilities presented by combining Java applets with EPIC, to increase the `intelligence' available at the page.

4. Conclusion

It has been argued that the principles of structuring the environment can be used to hide unnecessary detail from the student in a learning environment. EPIC represents an application of this principle. Although we haven't carried out any controlled studies on its effect of learning, the anecdotal evidence we have received from questionnaires suggests that the support it provides is helping.

Acknowledgements

Work on the EPIC system and courses has involved a number of staff at EPCC, in particular Alistair Ewing and Mario Antonioletti whose efforts have been invaluable.

References

[Kirsh95] David Kirsh. "Intelligent use of Space". Artificial Intelligence 73 (1995) pp31-68.

[Reason82] J.T. Reason and K. Mycielska, "Absent-Minded? The Psychology of Mental Lapses in Everyday Errors" Prentice-Hall, Englewood Cliffs, NJ, 1982.

[Reason90] J.T. Reason "Human Error". Cambridge University Press, Cambridge, England, (1990).

[Gibson77] J. J. Gibson. "The Theory of Affordances", in: R. E. Shaw and J. Bransford eds. "Perceiving, Acting and Knowing" Lawrence Erlbaum, Hillsdale NJ, 1977. Note: this paper is currently under consideration for WebNet'96

Digital Libraries Based on Full-Text Retrieval

Ian H. Witten, Craig G. Nevill-Manning and Sally Jo Cunningham
Department of Computer Science, University of Waikato,
Hamilton, New Zealand.
{ihw, cgn, sallyjo}@cs.waikato.ac.nz

Abstract: Because digital libraries are expensive to create and maintain, Internet analogs of public libraries-reliable, quality, community services-have only recently begun to appear. A serious obstacle to their creation is the provision of appropriate cataloguing information. Without a database of titles, authors and subjects, it is hard to offer the searching and browsing facilities normally available in physical libraries. Full-text retrieval provides a way of approximating these services without a concomitant investment of human resources. This presentation will discuss the indexing, collection and maintenance processes, and the retrieval interface, to public digital libraries.

Creating a publicly-available digital library presents some interesting challenges, particularly if the library is intended for serious professional work rather than casual browsing. First, the raw material: the collection must comprise text that can be placed in the public domain. Second, the selection of material: although a huge amount of public-domain text is present on the Internet, its quality is extremely uneven and only a tiny fraction is appropriate for inclusion in a library. Third, the format of material: it is helpful if the collection resembles the traditional form of typeset pages rather than raw electronic text. Fourth, cataloguing: appropriate bibliographic information in a usable format may be difficult to find and onerous to provide manually. Fifth, information retrieval: a uniform, easy-to-use, publicly-accessible interface is necessary.

There are two essentially opposite approaches to solving these problems. One is to arrange for authors, or their representatives, to contribute information to the library, so that a traditional author/title/keyword catalogue can be built through which readers access the collection. Because authors have a vested interest in seeing their work fully represented and properly catalogued, they will presumably be prepared to invest time and effort to ensure that full information is supplied. An extreme example of this philosophy is the Hypatia project (http://hypatia.dcs.qmw.ac.uk) which is creating a database of people, research interests, bibliography entries, and papers, on the basis that people offer information by subscribing individually and by academic department.

The second approach, which is pursued in this paper, is to gather material from public repositories without any active participation by authors or their institutions. This has the potential to create a vastly larger collection, but requires a new mechanism for searching the library because conventional bibliographic information is not available. An extreme example of this philosophy is the well-known AltaVista search engine which provides a full-text index to a vast universe of indiscriminately-garnered Web pages (http://altavista.digital.com).

Although the second approach provides the basis for the search engines that are rapidly becoming our primary access point to the Web, few digital library projects have adopted this philosophy. One that has is the New Zealand Digital Libraries project. This is unique in several respects. First, it provides a full-text index of the entire contents of each document, whereas other digital libraries index on user-supplied document descriptions, abstracts, or other document surrogates. Second, it makes a minimum of assumptions about conventions adopted by document repositories. Other schemes rely for their information on the index file that is present by convention in most directories of technical reports, or on other information provided explicitly for indexing purposes. Third, the work directly addresses the problem of building the library in a geographically remote location with high Internet costs-an environment in which the benefits of networked library technology are especially striking. Finally, the scheme is extremely economical in local disk resources.

Like many other digital library projects, we have chosen the domain of computer science technical reports to exemplify our approach. Much high-quality information already exists in digital form and is freely accessible on the Internet, and because the time value of information is high, computer science already relies more than most on pre-publication in the form of technical reports.

This paper begins by briefly surveying both Internet technical report collections and Web search engines. These exemplify the two different philosophies to collection-building outlined above. We then go on to describe our prototype digital library. We describe the indexing and retrieval mechanism, including the full-text index, and summarize the types of search provided, the indexes that are needed to support them, and the query interface through which they are made accessible. We go on to discuss the process of building the collection, and the crucial issue of what information is stored centrally and what can remain distributed. We then briefly summarize how the information is gathered, consider the factors that limit the size of the library, and review future directions for networked digital libraries.

Internet libraries for technical reports

Name	Domain	Reference	Sites	Documents	Indexed on
e-print archive	Physics	[Ginsparg 1994]		several million	user-supplied title, author, abstract
ntrs	NASA	[Nelson *et al.* 1994]		3 000 000	abstracts
harvest	CS	[Bowman *et al.* 1994]	380	37 000	limited info from several file types
ucstri	CS	[Van Heyningen 1994]	185	14260	file of abstracts at ftp site
dienst	CS	[Davis & Lagoze 1994]	15	-	bibliographic information provided
waters	CS	[Maly *et al.* 1994]	15	2 000	citation and possibly abstract
ncstrl	CS	www.ncstrl.org	55	-	bibliographic information provided
Karlsruhe	CS bib	liinwww.ira.uka.de/bibliography		560 000	full text of bibliographic entry

Table 1: Technical report indexes on the Internet

Several technical report searching and indexing systems have been developed in the past; some are summarized in [Tab. 1]. The physics e-print archive, operating since 1991, has already supplanted journals and pre-print mailings as the primary information dissemination point for several areas of physics. Documents are submitted by e-mail in TeX format, along with bibliographic information which is used for indexing. NASA has a large index to publications since 1962, but most references are to documents that are only available in physical form. Computer Science is well endowed with indexes. The HARVEST system contains a large number of documents, many of which are Web pages rather than technical reports. The documents are indexed on limited information extracted from the original files. UCSTRI allows searching of files of abstracts, which often appear in technical report FTP archives. While this information is limited in scope, the collection includes 185 sites and 14,000 documents. The DIENST and WATERS systems have been superseded by NCSTRL (see below). DIENST is a distributed library architecture, developed in a project involving several universities, that originally accommodated scanned images and indexed on text produced by OCR. WATERS included technical reports from several universities and provides bibliographic searches based on information provided by participating departments.

Recently, the NCSTRL (for "Networked Computer Science Technical Reports Library") project has arisen as a development of the DIENST and WATERS systems. NCSTRL is a collection of computer science technical reports made available for non-commercial and educational use. There are two levels at which organizations can participate. "Standard" participants run indexing, searching, and user interface software locally, whereas "Lite" participants maintain an ordinary FTP site of technical reports and supply bibliographic information to a special server. Lite connection is easiest but Standard connection allows a site to maintain its own search engines, to provide access to documents in multiple formats, and to provide customized user interfaces-thumbnail browsing, page zooming, page selection, support for logical document structure, etc. NCSTRL currently (July 1996) has 55 participants, of which nearly two-thirds are Lite.

Standard NCSTRL sites use the DIENST software, an open, distributed, implementation of digital library servers. DIENST provides three basic services at each site: a repository, indexes and a search engine, and a user interface. The repository can hold documents in different formats (PostScript, TIFF, GIF, HTML) and can handle the same document in multiple formats. Sites can use home-grown software for any of these components. For example, they can supply their own browsers (and some do). DIENST makes a distributed library appear as a single collection at the top level, even though the repositories, search engines, and browsing interfaces are distributed.

Web search engines

At the other end of the spectrum from contributory technical report libraries are search engines that seek and index Web pages. Typically, they permit a full-text search on the contents, or partial contents, of Web pages and some of the files that they point to-although most non-HTML file types, including PostScript, are excluded.

It is interesting to trace the historical development of such systems. In the pre-Web era, a system called VERONICA claimed virtually complete coverage of the documents pointed to by the GOPHER menu structure-although no attempt was made to index the document text. Early Web indexes such as ALIWEB were based on specially formatted submissions from service providers. JUMPSTATION provided three separate indexes to HTML pages: one for titles, a second for headings, and the third for the most frequent words in the text. The LYCOS system (www.lycos.com), at least in its original version, restricted the text searched to the first few lines of each page, and certain high-information-content keywords that appear later in the page. By 1995 LYCOS had encountered several million pointers to Web pages, but 75% of these were unexplored; at that time new documents were being indexed at the rate of 15,000 per day.

Although these early Web indexes were academic efforts, most present-day search engines are either funded by advertising or created by commercial firms that want to advertise their services and products in the information retrieval area. ALTAVISTA, for example, is a high-profile operation of Digital Equipment Corporation, and is exceptionally well-resourced to ensure that it provides impressive performance. Other search engines offer advertising, and in some cases it has been possible for Web page owners to pay to ensure that their pages figure prominently in the information retrieved by searches. Thus the indexes are competitive in nature; consequently all claim to provide the best service and it is very difficult to say with any accuracy how many pages they index-all tend to mention figures of around 50 million Web pages.

INFOSEEK (www.infoseek.com) and WEBCRAWLER (webcrawler.com) are commercial systems that resemble LYCOS, although at least at the beginning they offered considerably less coverage. OPENTEXT (www.opentext.com) is a more recent addition to the field. It offers ranked, Boolean and proximity searches that can be confined to summaries, titles, first headings, links, and body text. A year ago its index covered around one million Web pages, and around 50,000 pages per day were being added. ALTAVISTA is another, more recent, large index. It offers a basic ranked search facility in which one can additionally specify that certain terms are mandatory and others are prohibited. Terms can take the form of words, phrases and partial words. An advanced search facility allows Boolean expressions of search terms as well, separating the query specification from the way in which the results are to be listed. Finally, the EXCITE search engine site (www.excite.com) is eager to announce that its index is 50% bigger than its nearest competitor, that its indexing technology produces more relevant hits, and that its speed is unmatched by other search facilities.

The NZ digital library for computer science research

The New Zealand Digital Library for Computer Science Research is essentially an application of Web indexing technology to a digital library system. It incorporates features of both approaches, combining the unobtrusiveness of Web search engines with the selectivity afforded by confining search to a limited set of information archives. It currently provides access to 22,000 research documents on 250 sites worldwide (550,000 pages, 220 million words). This corresponds to 17 Gb of PostScript files containing technical reports in the computer science area; some relevant statistics are given in the upper block of [Tab. 2]. The library went on-line in May 1995 with an initial collection of approximately 2000 documents. Although until recently it has been publicized only within New Zealand universities, the library has been accessed from over 700 sites worldwide. The current implementation is clearly both useful and seeing use, and user feedback is overwhelmingly positive.

Indexing and retrieval

The key innovation in this project is that it dispenses with the traditional library catalogue, which relies on information about each item in the collection being supplied along with the item itself. If information providers must supply such information, the scope of the collection is thereby restricted. The alternative, cataloguing by library staff, means that the collection is limited by human resources. It was resolved that although the collection would be formed entirely from publicly-available repositories of text, the system should not require any effort on the part of participating repositories; indeed there should be no need for these information providers to be aware of their inclusion in the index. No special software, archive organizations, or file formats are required of the providers. The only information used for cataloguing is derived from the documents themselves.

		total	*per report*
Source documents	**Sites**	250	
	Documents	22 000 (550 million words)	10 000 words
		550 000 pages	25 pages
	Original PostScript	17 Gb	800 kb
	compressed with gzip	*6 Gb*	*270 kb*
Extraction process	**Text**	1.3 Gb	60 kb
	compressed with gzip	*430 Mb*	*20 kb*
	First page facsimiles	29 000 pages	1.5 pages
		300 Mb	14 kb
Indexed collection	**Text** **Index** **First-page images**	375 Mb 300 Mb 305 Mb	
	Total	980 Mb	

Table 2: Current size of the New Zealand digital library for computer science research

Full-text index

These considerations led to the use of a full-text index instead of a library catalog as the main retrieval mechanism. The index makes the entire text of all documents available for retrieval, rather than some much more restricted list of keywords as is the case in many computer-based text retrieval systems. From the point of view of building the system, this has the advantage that nothing more is needed to construct the index than the plain text of the documents in the library: it is a way of finessing the usual requirement for traditional bibliographic database information such as author, title, publisher, and so on. From the point of view of the user, it provides a powerful tool for searching for information: we consider below whether this can supplant the more traditional forms of library access-by author, title, date, subject, and so on.

The search engine for the library is the public-domain system MG. Tailored for highly efficient storage of full-text databases, MG can pack an index to a large collection of text into only 5% of the size of the original text [Witten *et al.* 1994]. Further, it responds rapidly to queries: experiments with the 750,000 document TREC collection produce ranked output for queries of forty to fifty terms within three to five seconds.

Types of search

MG supports the usual Boolean and ranked keyword searches over the full text of the document. In the digital library, different kinds of search are implemented as follows.

Author/title. In most reports, the first page gives bibliographic information such as title and author. By limiting attention to this page, the user can approximate a search based on such information. For example, an initial page search for documents authored by *Knuth* will not retrieve documents that merely cite his work.

Publication date. Most reports also include publication date on the initial page, and the same technique serves for publication date searches. Alternatively, this facility could be simulated by permitting the user to search on the date in which a technical report was entered into its repository, although such a strategy is likely to produce uneven results because some reports are placed in repositories long after they were originally produced.

Page searching. The digital library stores the full text of technical reports, and supports searching over the complete document text. This is very useful in performing very general searches with high recall, that is, ones that retrieve a high proportion of the relevant documents in the collection. However, a large number of irrelevant documents ("false drops") can be expected as well. For example, in a search for *information retrieval*, many technical reports will contain these terms somewhere in the document, perhaps widely separated. The greater the physical distance between the two words, the less likely that the document is about the subject *information retrieval.* A limited kind of proximity searching is supported by page-level indices. which require the query terms to appear on the same physical page of the document. A still higher degree of precision can be achieved by specifying an exact co-location of words in a phrase. Phrase searching is implemented within MG by post-processing query results; a string search for phrases can be performed on documents returned by any query.

Case folding/truncation/exact match. Retrieval systems commonly allow users to decide whether or not query terms should be exactly matched in the document returned. Exact matches are useful, if not crucial, in constructing some types of search: for example, in locating the author *Gray* rather than the color *gray*, or for finding documents about *NeXT* computers or the *SMART* system. On the other hand, case folding helps to avoid artificial distinctions between capitalized and non-capitalized forms of the same word (such as *Information* and *information*) or between different forms of a word (*retrieval, retrieve, retrieving*). These truncated terms of the root word should be sought automatically, without forcing the user to list all alternatives. The digital library allows the user to alternate between these types of search by selecting either exact match or case folding/stemming.

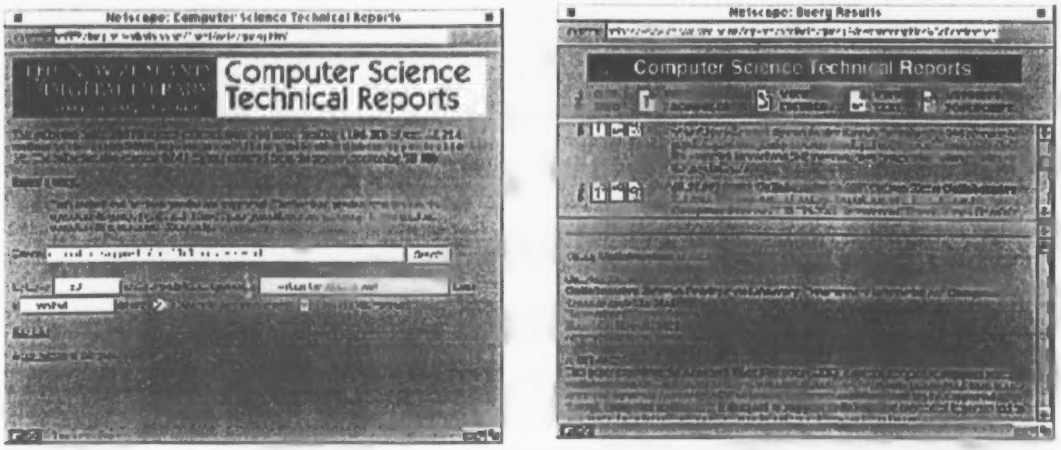

Netscape: Computer Science Technical Reports
Computer Science Technical Reports
Netscape: Query Results
Computer Science Technical Reports

Figure 1: Interface to the digital library: the query page and a typical response

Query interface

The World Wide Web is the access medium for the digital library. [Fig. 1] shows the query page and a typical response. Because of the Web's page-oriented nature, the document list, query options, and returned documents are placed on separate pages. The maximum number of matches returned for a ranked query is a potential transmission bottleneck, for it is easy to issue a query that returns many thousands of documents. This is initially set to 40, which gives good performance over the Internet, but can be increased by the user. Because WWW transactions are stateless and the user is unknown to the server, it is impossible to provide services such as remembering queries, displaying a history, or saving profiles, without requiring users to identify themselves explicitly.

It is clearly necessary to share the query engine between users. MG was not designed to deal with simultaneous multiuser access, and while it is possible to launch a process for each query, the startup overhead makes this very inefficient. Consequently we have arranged for access to one persistent copy of the query engine to be shared between incoming queries. One copy runs continually, and additional software provides exclusive access for the duration of each query.

Building the collection

The lowest common denominator for representing information in conventional libraries is paper, and many digital library efforts involve scanned paper documents (e.g. [Crocca & Anderson 1995; Van House 1995]). In the world of electronic information, PostScript, rather than plain, unformatted text, is the closest analog to paper as a document storage medium, and its page-based nature turns out to be helpful in structuring the collection. In order to build the index, it is necessary to be able to extract plain, unformatted, text automatically from the documents. In fact, the system design accommodates not just PostScript but any format from which paginated ASCII text can be extracted-for example DVI, RTF or HTML files. Document images can be accommodated by OCR-ing them for indexing purposes: the inevitable recognition errors will reduce the quality of the index, but this can be ameliorated by using ranked queries containing redundant terms. However, from an initial investigation it appeared that PostScript files are almost universal in computer science technical report archives, and the system currently deals only with this format.

Archives of technical reports can be located through several lists maintained on the Internet, and recursively descending the directory hierarchy looking for (possibly compressed) PostScript files. Each file is downloaded, along with its size and date, and the appropriate information is extracted.

BEST COPY AVAILABLE

File Name : nzdl-ps-extraction.eps
Title :
Creator : Diagram
CreationDate : Mon Aug 21 14:30:03 1995
Pages : 0 0

Figure 2: Conversion from PostScript: a PostScript file, the text extracted from it, and a facsimile image

Information stored centrally

With an Internet-based digital library, a crucial question is how much information to store centrally. It was decided that the library would comprise an index and search engine, and the documents themselves would remain in their original repositories. Periodically, all repositories are scanned to refresh the index, adding any new documents and removing references to those that have been deleted. There is no intention to provide an archiving service that will retain documents that are removed from their original sites.

In addition to the index, a facsimile image of each document's first page or two is retained locally so that users can read the title and abstract, and sample the look and feel of the original. The plain, unformatted, text must be extracted from the documents to build the index, and it proves expedient to retain a full copy of this text locally as well. This is useful in its own right for browsing: users can examine the text without going to the trouble of downloading PostScript. Moreover, the text of the collection provides an excellent foundation for bibliometric research. [Fig. 2] illustrates (at the top) an original document in PostScript form, and (at the bottom) the two files extracted from it: on the left the full ASCII text, and on the right a facsimile image of the first page.

Text extraction

Although the words of a technical report usually appear as plain text within a PostScript file, they are thoroughly intermixed with PostScript language commands and internal data. Words appear within parentheses, but so does internal information such as font names and error messages. Spaces are not explicit, but are coded implicitly in terms of the placement of words on the page. Finally, whole words are not always bracketed together: to give greater control over spacing, word fragments are often placed individually. For these reasons, text cannot be extracted reliably by syntactic analysis of the PostScript file. [Fig. 2] illustrates some of these problems.

The solution is to prepend a PostScript prologue that redefines the operators responsible for placing text on the page, and execute the resulting file using a PostScript interpreter. The redefined operators write the text fragments to a file, taking note of their position on the page in order to insert spaces and new lines where appropriate. This technique is based on part of the GhostScript distribution (the *ps2ascii* program), but is simpler and far more robust in terms of the variety of PostScript files it can handle. In fact, *ps2ascii* failed on 75% of 144 reports chosen at random (each from a different site), while another converter, that used in the Harvest system [Bowman *et al.* 1994], failed on 40%.

`/show {`	redefine the show operator to be:
`currentpoint pop`	get the current x coordinate
`X sub 5 gt { ( ) print } if`	output a space if x has increased by at least 5 points
`dup print`	output the string
`systemdict /show get exec`	execute the original show to update the current point
`currentpoint pop /X exch def`	record the new x position
`} def`	

Figure 3: The prologue used to extract text from PostScript documents

[Fig. 3] shows the a simplified version of the PostScript prologue used for this process. It extracts not only the body of the report, but also any text in figures and tables. Although an absolute specification of the minimum inter-word spacing (5 points) is shown, the actual implementation sets this threshold by analyzing the report itself. Moreover, although for indexing purposes only the stream of words needs to be extracted, the Digital Library provides a HTML approximation to the original document for browsing purposes. We have developed simple heuristics to recognize paragraph boundaries, and to solve other practical problems with the output of the extraction process-such as dealing with non-ASCII characters, dehyphenation, and page reversal in case the PostScript file holds the document back to front (these situations are not easy to detect).

First-page facsimile

The facsimile image of the first page shows the user the beginning of the document as it actually appears, complementing the raw text of the full paper. This provides an analogous service to a library supplying users with fax copies of the first page or two of a document.

One approach is to store the PostScript version of the first page. However, PostScript documents invariably contain a lengthy prologue that defines fonts and special functions. This prologue is essential to interpret the first page, and it turns out that in most cases the size of the file required to reproduce the first page is nearly as large as the entire document. Consequently the facsimile is stored as a screen-resolution bitmap, so that the file size is bounded by the size of a standard page. The facsimile images are produced by a PostScript interpreter as a 75 dpi bitmap, and are saved as graphics files in GIF format.

The first page of many technical reports shows little more than an institutional logo along with the title and authors' names. The second page may also give little information. In order to ensure that a page with some content is imaged, the information in a page is estimated by the size of the GIF file that represents it. Because these images are compressed, pages with large areas of white space and big letters compress well, to a few kb, while pages full of actual text occupy over 10 kb. The extraction system renders pages in sequence until the sum of the file sizes exceeds this figure, a simple heuristic that seems to work well in practice.

Collection maintenance

All that is necessary to maintain the library's collection is to ensure that the documents indexed continue to exist. This can be checked by periodically examining the technical report repositories for changes and updating the collection accordingly. Considerable benefit is obtained by not relying on any cataloguing information stored with the repository itself: for example, frequent maintenance problems in technical report servers are caused by changes in the bibliography file format, and inconsistencies in the information, in the repositories that they index [Van Heyningen 1994].

One source of growth for the collection is when new documents in known repositories are located during a routine examination of currently indexed sites. New sites can be detected by various means: monitoring standard Internet lists for new additions; manually scanning the newsgroups that announce them; and encouraging users to email suggestions to a central coordinator.

Gathering the Information

One of the motivations for a New Zealand digital library is efficient use of the expensive transpacific Internet link. Computer scientists can search for and preview technical reports locally before downloading the full document file, thus encouraging exploration without concern for network charges. However, to build the full-text index it is necessary to examine the contents of each report. Transmitting all of them across the Pacific would negate any cost benefits the project might offer.

Consequently a distributed scheme is used to create the index. In the first stage, a computer in North America downloads each technical report, extracts the facsimile images and raw text, sends them to New Zealand, and deletes the report. This process is illustrated in [Fig. 4].

This operation significantly reduces transmission cost on the Pacific link. The 17 Gb of uncompressed PostScript currently indexed by the library is actually stored in repositories in compressed form, occupying approximately 6 Gb. In this form, each report takes about 13 seconds to download to the North American site, whereas it would take have taken about 36 seconds to download to New Zealand because of the low bandwidth of the Pacific link. In actuality, only 1300 Mb of uncompressed text, which reduces to 430 Mb when compressed, need be transmitted, which takes about 4 seconds per report. Added to this is the cost of transmitting facsimile images: 305 Mb in total, or another 4 seconds per report.

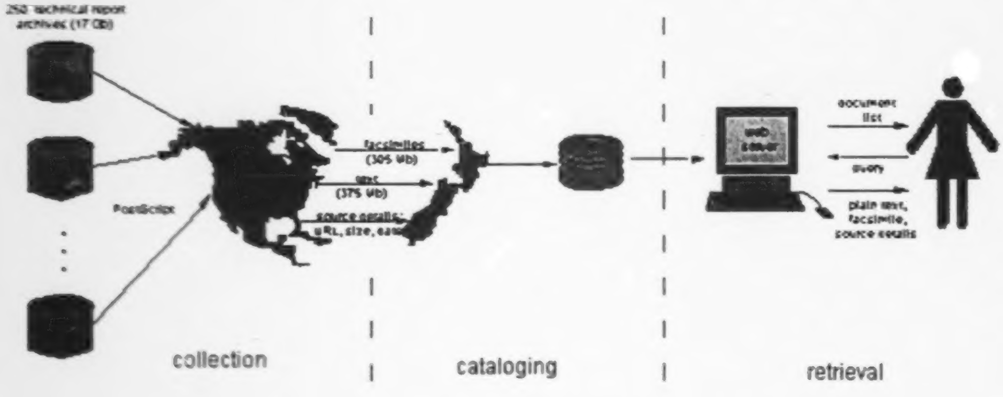

250 technical report archives (17 Gb)
PostScript
facsimiles (305 Mb)
text (375 Mb)
source details: URL, size, date
web server
document list
query
plain text, facsimile, source details
collection
cataloging
retrieval

Figure 4: Architecture of the digital library

Size of the library

The lower block of [Tab. 2] summarizes the size of the document collection and the storage requirements for the MG indexes that comprise the library. The total space occupied by all indexes is

BEST COPY AVAILABLE

300 Mb, a mere 2% of the total size of the PostScript files that are indexed. The text of the technical reports occupies 375 Mb when compressed by MG. Finally, the first-page facsimiles occupy an additional 305 Mb, giving a total disk storage requirement of just under 1 Gb, 6% of the total PostScript size.

What are the limits to the size of a library of this nature, using current technology? All sizes involved increase linearly with the volume of text, and a factor of ten growth would only involve a few Gb of disk storage. Retrieval time is virtually independent of database size: it takes two disk seeks per query term and two per document retrieved. The database inversion process is more likely to be a limiting factor. An improved algorithm has recently been reported that is faster and requires less main memory than MG: extrapolating from experiments with a 2 Gb text collection it is estimated that 5 Gb of raw text can be inverted in twelve hours with only 40 Mb of main memory [Moffat & Bell 1995]. This corresponds to growth of the digital library by a factor of four. It provides an indication of what is certainly possible, although there is no reason to believe that it represents a limit. The resulting collection would index 80,000 documents comprising two million pages or 45 Gb of PostScript, and the size of the centralized information, including plain text, indexes, and first-page facsimiles, would amount to just 4.5 Gb.

Nevertheless, an all-inclusive information collection policy is basically unscalable and will become infeasible as the number and size of technical report repositories grows. There are several possibilities for culling the collection when that becomes necessary. One is to monitor participating users' access, and note the repositories that receive the least use. This regulates the growth of the collection by the size, diversity, and level of activity of the user population rather than by the growth of the bibliographic universe. Another is to use the characteristics of the documents themselves to determine what to cull from the collection. For example, citation studies could identify documents that do not appear to contribute to the literature.

Future Directions

The operation of the digital library hinges on the idea of extracting information automatically from PostScript. Presently, words and page boundaries are extracted, and the process is entirely algorithmic. The collection would be enhanced if reliable ways could be devised for determining bibliographic details of the reports so that a proper catalogue could be built, rather than relying on their appearance on the first page (which may be a cover page). Moreover, the text collection provides an ideal basis for bibliographic studies in computer science. If it were possible to locate bibliographic reference sections within documents, a separate database could be constructed that would be extremely useful for citation analysis. Hence we need to consider how to extend the extraction process by incorporating heuristics to identify some document structure.

Users find it hard to choose between document- and page-level retrieval because it is not clear what the basis of the choice should be. It may be that the page-level index suffices for all retrieval, and the document-level one should be abandoned. Perhaps the problem can be solved by a more sophisticated information display mechanism: for example, TileBars [Hearst 1995] provide a way to visualize the distribution of query terms throughout a document that should be easy to integrate into our system. A number of other searching mechanisms would be very useful. Mixed Boolean and ranked queries would allow users greater control over the documents that are retrieved. Browsing by location needs some human pre-processing to identify machine locations by institution rather than by network address. Even simpler, and also useful, would be to allow users to look at the directory that a particular report comes from, because sometimes multipart documents are stored as separate files. Finally, co-location queries allow users to examine which words occur together; this is useful for textual analysis purposes.

Conclusion

We have distinguished two opposing philosophies for the construction of networked digital libraries. One is to have contributors supply bibliographic details for the information that they place in the collection. The other is to garner the collection automatically and use a full-text index for access instead of an ordinary library catalogue.

This paper has focused on the second approach. With it, public digital libraries can be constructed on the Internet entirely automatically, in any area for which repositories of suitable text can be located. Extracting all index information from the documents themselves is feasible if full-text indexing is used, and eliminates the manual cataloguing effort involved in creating and maintaining the library. The material in such a library can be distributed globally. The central database in our prototype library, which comprises a full-text index, facsimile images of the first page or two of each document, and the plain text of all documents, represents only 10% of the collection size.

Looking into the future, it is likely that new generations of digital library will marry the two approaches. The first offers high-quality cataloguing information, while the second provides significantly increased coverage. Improved techniques for information extraction from text, along with large public-domain bibliographies, offer the possibility of being able to match reports in a collection with items in a bibliography file, thus providing cataloguing information at no additional cost. Moreover, personal contributions to people-oriented research databases like Hypatia are likely to provide more authoritative reference information than do general bibliographies, so bibliographic quality can be increased by amalgamating information from different sources, tagged according to likely reliability.

Finally there is the problem of distribution. Architectures like HARVEST and NCSTRL provide a distributed infrastructure that underpin all aspects of the collection. However, there is a danger in being *too* distributed: whereas users want to see a unified system, these schemes allow sites to provide their own browsing software through which their repository must be viewed, and this nonuniformity can be a great annoyance in practice. Web search engines, in contrast, are not distributed, though they may involve multiprocessors accessing the same database in order to provide adequate power. Again we will probably see an amalgamation of the two approaches, and indeed distributed indexes expressly designed for multi-collection environments are a current research topic in the information retrieval community.

Digital libraries represent one way of dealing with the new reality of Internet publishing. Making a minimum of assumptions, a library based on full-text retrieval imposes structure on a fundamentally anarchic, uncatalogued, system, giving information consumers a tool to find what they need.

References

[Bowman *et al*. 1994] Bowman, C.M., Danzig, P.B., Manber, U., and Schwartz, M.F. (1994) "Scalable Internet resource discovery: Research problems and approaches," *Communications of the ACM 37*(8), pp. 98-107.

[Bowman *et al*. 1994] Bowman, C.M., Danzig, P.B., Manber, U., and Schwartz, M.F. (1994) Scalable Internet resource discovery: Research problems and approaches. Communications of the ACM, 37 (8), 98-107.

[Crocca & Anderson 1995] Crocca, W.T. and Anderson, W.L. (1995) Delivering Technology for digital libraries: experiences as vendors. Proc. Digital Libraries '95, 1995, Austin, TX. 1-8.

[Davis, & Lagoze 1994] Davis, J. & Lagoze, C. (1994) "A protocol and server for a distributed digital technical report library," Technical Report 94-1418, Computer Science Department, Cornell University.

[Ginsparg 1994] Ginsparg, P. (1994) First steps towards electronic research communication, *Computers in Physics 8*(4), pp. 390-401.

[Hearst 1995] Hearst, M.A. (1995) TileBars: visualization of term distribution information in full text information access. Proc. CHI'95, 56-66.

[Maly *et al*. 1994] Maly, K., Fox, E.A., French, J.C., and Selman, A.L. (1994) "Wide area technical report server," Technical Report, Dept. of Computer Science, Old Dominion University.

[Moffat & Bell 1995] Moffat, A. and Bell, T.A.H. (1995) In situ generation of compressed inverted files. J American Society for Information Science, 46 (7), 537-550.

[Nelson *et al.* 1994] Nelson, M.L., Gottlich, G.L., and Bianco, D.J. (1994) "World Wide Web implementation of the Langley Technical Report Server," *NASA Technical Memorandum 109162*, Hampton, Virginia.

[Van Heyningen 1994] Van Heyningen, M. (1994) The Unified Computer Science Technical Report Index: Lessons in indexing diverse resources. Proc. Second International WWW Conference, Chicago.

[Van House 1995] Van House, N.A. (1995) User needs assessment and evaluation for the UC Berkeley electronic environmental library project: a preliminary report. Proc. Digital Libraries '95, Austin, TX. 71-76.

[Witten *et al.* 1994] Witten, I.H., Moffat, A., and Bell, T.C. (1994) Managing Gigabytes: Compressing and indexing documents and images. New York, Van Nostrand Reinhold.

Acknowledgments:

Thanks to Todd Reed, Mike Vallabh and Brent Summers for technical assistance, and to Tim Bell and Mark Apperley for encouragement and advice. Bruce McKenzie provided the prototype WWW interface to MG, and the Computer Science Department at Calgary generously served as our North American site. This research was supported by the NZ Foundation for Research, Science and Technology and by the NZ Lotteries Grants Board.

Supporting Provision and Access of Educational Visual Resources on the Web

Zhengmai Zhao
Department of Media & IT, De Montfort University, UK
zmzhao@dmu.ac.uk

Jian Hu
Roslin Institute (Edinburgh), UK
jian.hu@bbsrc.ac.uk

Abstract: Our recent investigation of user requirements show that an easy and direct way for delivering educational visual materials through the Web is strongly demanded by lecturers teaching subjects related to Arts and Design, and that multiple approaches to retrieval of such kind of information is preferred by students. In this paper, we will present an information model to meet the user requirements along with its two implementations: datafile-based and database-based.

Introduction

Visual materials such as slides and video tapes are vital, particularly in the subject areas related to Arts and Design, to aid the student's understanding of the subject and to engender critical thinking. Normally lecturers who have a large collection of visual materials do not have a convenient way to make them accessible to the students. In most cases, they can be only selectively shown to students during lectures. Even though such materials are sometimes made available through libraries, the number of copies is usually too small to meet the demand, and lecturers may be reluctant to provide such materials in this manner due to the subsequent lack of control over their uses.

Recent advances in information technology especially the World-Wide Web (WWW or Web) have created an exciting opportunity to substantially change this situation. Among its tremendous potentials in a wide variety of areas, WWW is an excellent distribution channel for educational visual resources due to its multimedia/hypermedia ability, platform-independence, and world wide access [Berners-Lee et al. 1994]. Although WWW provides an attractive platform for delivery of educational visual resources, it does not necessarily lead to a successful educational application. Without having a proper information model and a range of tools to support both provision and access of such resources on the Web, the medium may fail.

Our investigation of user requirements has shown:

- Lecturers prefer a direct control over their own visual resources, i.e., the so-called federal manner, than a centralised one, e.g., an imagebank which is created and maintained through a system administrator rather than the resource owners.
- Lecturers, especially those lacking computing experience, demand a browser-based interface which will allow them to create and modify their on-line visual resources easily and without having to learn many technical skills.
- Students prefer more approaches to retrieval of information than just keyword search. This conforms with the research result [Furnas et al. 1983; Gomez et al. 1990] which indicates that keyword search is not always efficient because the user might not know what he/she is exactly looking for or/and because the user may use different words or terms than those used by the

information provider. The possibility to be able to browse visual resources is strongly demanded as an equally important approach for information retrieval along with the keyword search.

Bearing the user requirements in mind, we investigated a number of projects including ELISE (Electronic Library Image Server for Europe) [Eyre 1994], STILE (Students' and Teachers' Integrated Learning Environment) [Ruggles et al. 1995; Zhao & Patel 1995; Zhao et al. 1996], and the Network Delivery of Multimedia Resources to the Academic Community project (URL: http://WWW.ets.bris.ac.uk). However, no systems have been found which could largely satisfy our user's requirements. They either lack means to allow students to browse the visual resources or means to enable lecturers to supply the resources on the Web easily and in a federal manner. Therefore, the objective of the work reported in this paper is to reduce the gap so as to increase teachers' power to manage and deliver visual materials, and to increase students' power to discover and access visual resources relevant to their learning needs. The remainder of this paper will first provide a description of a system which aims to meet the user requirements listed above. Following that, an application of the system is described as a further illustration of the system functionality. Finally, we propose a new implementation of the system based upon a relational database management system - INGRES.

Description of the System

The system has two major features: allowing provision of visual resources on the Web easily and in a federal manner, and allowing access of visual resources on the Web by both querying and browsing. The system has a client/server structure based on the architecture of the Web. The server basically consists of a Web daemon (e.g., CERN server, NCSA httpd), raw visual resource bases, and a range of tools to support both provision and access of the visual resource. The visual resource bases can be owned and maintained by different providers. A provider can have one or more resource bases. The client basically includes a Web browser (e.g., NetScape, Microsoft Internet Explorer), and viewer applications where needed. It provides interfaces to the searching tool, browsing tool, and the editor. Of course, the editor is available only to the provider. The architecture of the system is illustrated in Figure 1, and each element of the system will be explained in more detail in the following paragraphs.

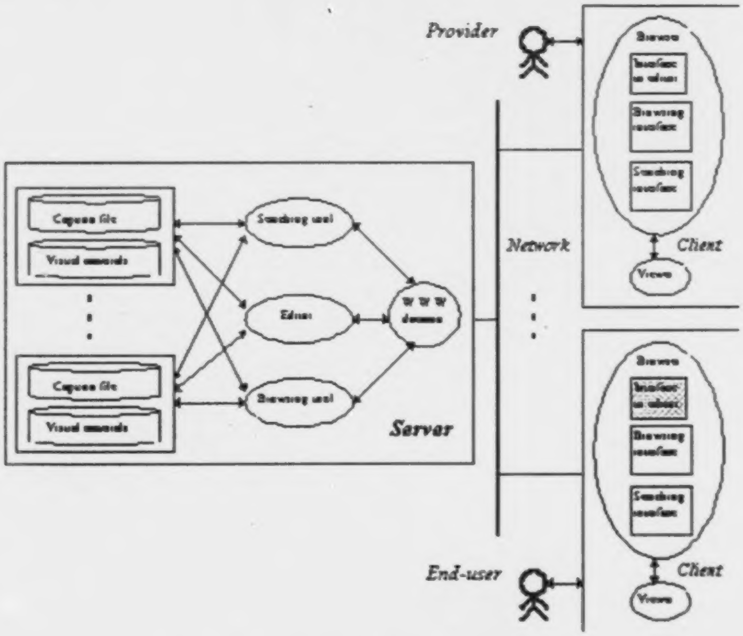

Provider
Browser
Interface to editor
Browsing interface
Searching interface
Caption file
Visual materials
Searching tool
Editor
W W W daemon
Browsing tool
Caption file
Visual materials
Server
Network
Client
Viewer
Browser
Interface to editor
Browsing interface
Searching interface
End-user
Client
Viewer

Figure 1: Architecture of the system.

A visual resource base comprises a collection of visual materials (e.g., images, video clips) and a caption file. The visual materials should be in the format the Web can handle, e.g., GIF, JPEG, MPEG. They

BEST COPY AVAILABLE

should be placed into the provider's Web directory tree so that they can be addressed as URLs. Textual information of the visual material is stored in the caption file as a set of records. The format of records is defined at the beginning of the file which includes attributes of each field such as title, length, type of content (text or HTML or URL), whether or not it is searchable, whether or not it will be displayed in query results. A collection of visual materials and its caption file are associated with each other via a unique field in the caption file, which holds a URL as its value.

The system provides three types of facilities to support information delivery and access, i.e., searching, browsing, and editing. They are all implemented by means of a combination of FORMS in the Web browser and CGI (Common Gateway Interface) scripts on the Web server. First of all, the editor enables the supplier to define a new caption file, and then to input data into and make changes when necessary to the file, directly through a Web browser. Secondly, the search facility[1] allows the end-user to make a query to the resource base from a Web browser. A query normally results in a list of keywords which represent items matching the query. The user can then choose any of them by clicking it to ask for its more detailed information which contains full textual explanation of the item as well as visual hints (e.g., thumbnails of images) as anchors of the eventual visual material. Finally, the browsing facility allows the end-user to browse a visual resource base according to the hierarchy of classification of the visual material. The classification hierarchy is in fact embedded within the caption file. Once a browsing session is initiated by a user, the browsing tool starts generating and displaying a partial hierarchy to the user step by step following his/her selection.

An Application of the System

The system is being used to build a personal on-line image collection about architectural history[2] , which will be introduced to students as resource material for courses related to the History of Architecture and Design. The collection will eventually contain a large number of photographs of structures from a number of different countries. Each photo is attached with a piece of textual information including name of structure, architect(s), location of structure, date when structure was built, type of structure, and its architectural style. In order to build the resource base, the first task was to define the format of the caption file via the browser-based editor. As a result of this, the system created a text file with the specification of how the textual information of images would be stored and treated. In this example, each record of the file comprises six fields corresponding to the items listed above, and a special field with the URL of an image as its value. Among the total seven fields, only the first six ones are searchable, and only the name and location of structure would be displayed in consequent query results as the key of a piece of structure. Having had defined the caption file, the next task was to input the data into the caption file by using the adding function of the editor. Prior to these tasks, the photos had been digitised and stored in the provider's Web directory tree as GIF images.

The appendix displays some screens from a searching session and a browsing session with the current architectural image collection. Screen 1-3 represent a searching session: a user made a query to the collection by entering in the searching interface "Manhattan, New York" and "Domestic" as the location and type of structure (Screen 1); ten items were found in the current collection and the keys of them were displayed to the user (Screen 2); and the user asked to see more about The Dakota Apartment by clicking on its name and its detail was shown with miniatures of its full photos (Screen 3). Screen 4-6, and 3 represent a short browsing session: a user started his/her trip with a list of classifications of structure (Screen 4); he/she clicked "Type of structure" for some reason and was displayed all names of types of structure (Screen 5); he/she was particularly interested in domestic structures so clicked on "Domestic" and all domestic structures were turned up with their key and a thumbnail of a representative photo (Screen 6); and he/she picked up a building for more detail, which happened to be the same one as the previous user chose (Screen 4).

A Database Implementation of the System

A database system consists of a software tool, called a Database Management System (DBMS) and one or more databases that it manages [Date 1990]. A DBMS provides facilities to describe all relevant data in an organisation at a relatively abstract level, or logical level. Application programs usually operate against the logical data structure while the details of physical data operation are insulated from users.

To allow users or application programs to operate upon the database, a DBMS provides some kinds of interfaces, such as command languages. The most widely used command language for the relational database is the Structured Query Language (SQL) which has a great generality and flexibility to allow applications to retrieve and update the database at a higher level, i.e. without any concern about the physical data manipulation. This is critical if the data structure is complex. Other important features of a DBMS include concurrent data access management, security and integrity control. They make it possible for multiple users to use the database concurrently and securely, so that each user is given the illusion of being able to use data on his/her own.

If a database system is used, it raises a demand for Web to database interfaces. A conventional Web to database interface is usually implemented as a CGI program which contains all data operations and HTML statements, e.g. a C/embedded-SQL program. Problems with this method are mainly related to the interface maintenance. Whenever the interface needs to be changed, the source code will have to be modified and recompiled. Another approach is to adopt a Web to database interface building tool which allows the application user to construct a Web interface through a set of user-customised forms written in an interface tool language. Such a tool language is mainly a combination of HTML statements and SQL-based data manipulation statements, which would be sufficient to describe interface displays and to specify database operations. The WebinTool is such a software package and more details about WebinTool can be found in [Hu et al. 1996].

A prototype with the same functionality described in the previous section but adopting a database approach has been implemented using the INGRES database system and the WebinTool. The system architecture of this approach is shown in Figure 2.

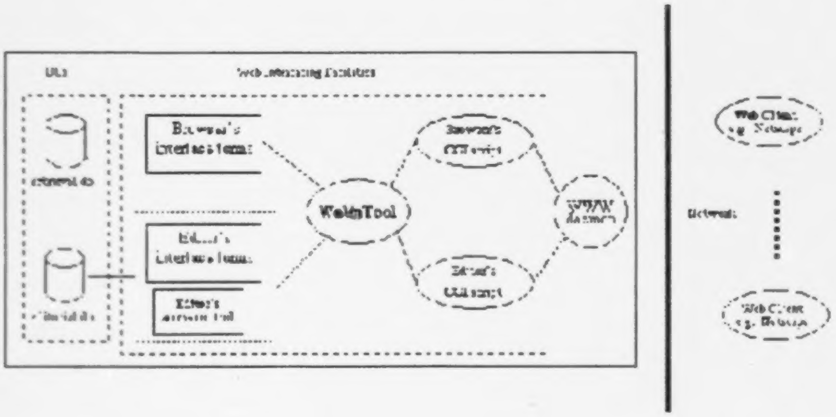


Figure 2: Architecture of the database-based system.

There are two copies of the database: the retrieval one for public use, and the editorial one for the database administrator and data supplier's use. The data in the editorial database is copied onto the retrieval database at a regular interval, e.g. at every night.

The Web interfacing facilities include:

- Web browser: It performs the searching and browsing functions based upon a set of WebinTool forms. A simple CGI script specifies the location of these forms as well as the database environment.
- Web editor: It performs the editing work based upon a set of WebinTool forms, and a security file which specifies the access control. Similar to the browser, there is also an editor CGI script.

Our practice has shown that this framework can effectively and efficiently fulfil the system requirements:

complex data structure can be defined and complex queries (for browsing and searching) can be constructed based on SQL statements; data providers can use the Web interface to manipulate their data conveniently and securely with very little learning; and interfaces can be rapidly built up and easily maintained.

References

[Berners-Lee et al. 1994] Berners-Lee, T., Cailliau, R.; Luotonen, A. Nielsen, F. & Secret, A. (1994). The World-Wide Web. Communications of the ACM, 37(8), 76-82.

[Date 1990] Date, C. J. (1990). An Introduction to Database systems (fifth Edition), Vol. I. Addison-Wesley.

[Eyre 1994] Eyre, J. (1994). The ELISE project: Visual information retrieval and delivery. In A. H. Helal & J. W. Weiss (Eds.), Information superhighway: the role of librarians, information scientists and intermediaries (pp 90-107). Essen: Universitats bibliothek.

[Furnas et al. 1983] Furnas, G., Landauer, T., Gomez, L. & Dumasi, S. (1983). Analysis of the potential performance of key-word information systems. Bell Systems Technical Journal, No. 62, 1753-1806.

[Gomez et al. 1990] Gomez, T., Lochbanm, C. & Landauer, T. (1990). All the right works: Finding what you want as a function of richness of indexing vocabulary. Journal of the American Society for Information Sciences, No. 41, 547-559.

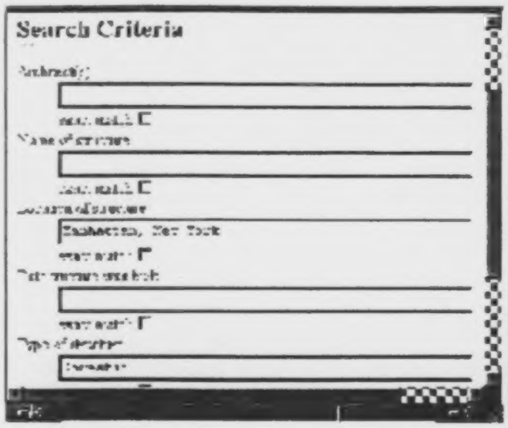
[Hu et al. 1996] Hu, J., Nichosolson, D., Mungall, C., Hillyard, A. L., & Archibald, A. L. (1996). WebinTool: a generic Web to database interface building tool. Proceedings of the 7th International Conference and Workshop on Database and Expert Systems Applications (DEXA '96), Zurich, Switzerland, September 1996.

[Ruggles et al. 1995] Ruggles, C., Underwood, J. & Walker, D. (1995). STILE: a hypermedia approach to a flexible, interdisciplinary resource base for teaching and learning. Innovations in Education and Training International, 32(3), 209-219.

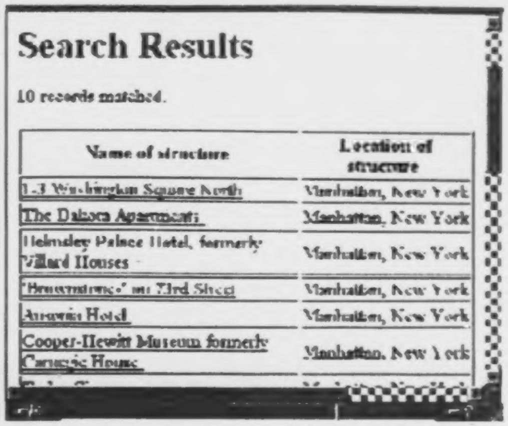
[Zhao et al. 1996] Zhao, Z., Cook, J., & Higgett, N. (1996). On-line learning for design students. ALT -J Journal, 4(1), 69-76.

[Zhao & Patel 1995] Zhao, Z., & Patel, D. (1995). An on-line resource-based learning environment. In D. Jonassen, & G. McCalla (Eds.), Proceedings of ICCE 95 - International Conference on Computers in Education (pp. 315-323). Singapore, 5-8 December, 1995.

Appendix



Search Criteria

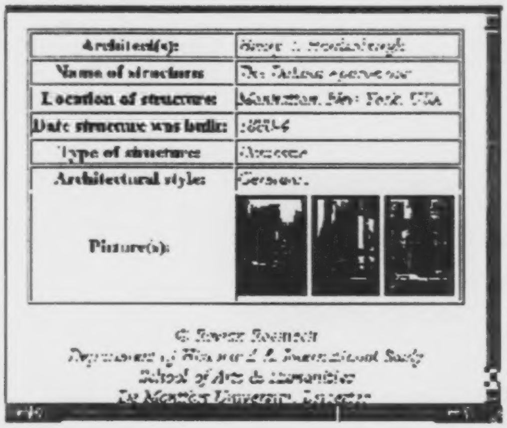


Search Results

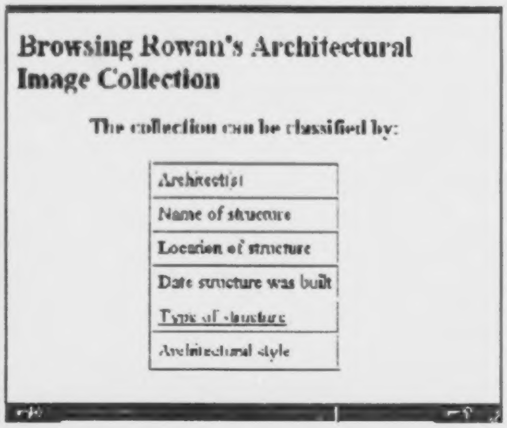
10 records matched.

Name of structure	Location of structure
1-3 Washington Square North	Manhattan, New York
The Dakota Apartments	Manhattan, New York
Helmsley Palace Hotel, formerly Villard Houses	Manhattan, New York
"Brownstones" on 73rd Street	Manhattan, New York
Ansonia Hotel	Manhattan, New York
Cooper-Hewitt Museum formerly Carnegie House	Manhattan, New York

Screen 1 Screen 2



Architect(s):	[illegible]
Name of structure:	The Dakota Apartments
Location of structure:	Manhattan, New York, USA
Date structure was built:	[illegible]
Type of structure:	[illegible]
Architectural style:	[illegible]
Picture(s):	

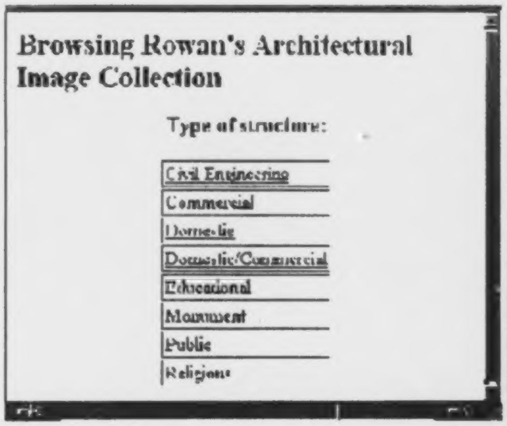


Browsing Rowan's Architectural Image Collection

The collection can be classified by:

- Architect(s)
- Name of structure
- Location of structure
- Date structure was built
- Type of structure
- Architectural style

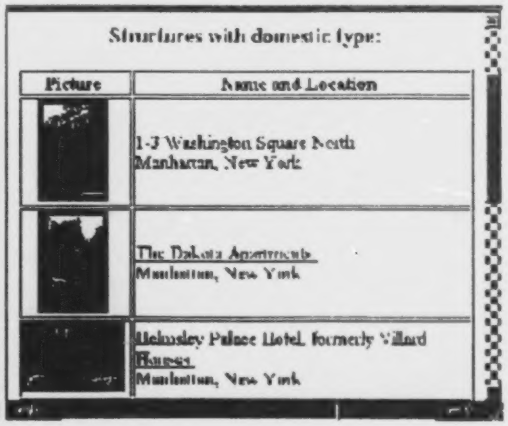
Screen 3 Screen 4



Browsing Rowan's Architectural Image Collection

Type of structure:

- Civil Engineering
- Commercial
- Domestic
- Domestic/Commercial
- Educational
- Monument
- Public
- Religious



Structures with domestic type:

Picture	Name and Location
	1-3 Washington Square North Manhattan, New York
	The Dakota Apartments Manhattan, New York
	Helmsley Palace Hotel, formerly Villard Houses Manhattan, New York

Screen 5 Screen 6

BEST COPY AVAILABLE

Pstohg - A Tool for Integrating Presentations into Hypermedia Systems

Jörn Bollmeyer
HEINZ NIXDORF INSTITUT
Universität-GH Paderborn, Fürstenallee 11, D-33102 Paderborn, Germany
email: bolle@uni-paderborn.de

Hypertext systems are becoming increasingly important in educational contexts where they are often used as repositories for teaching material such as textbooks or other documents related to lectures [cf. Norman 1990]. On the other hand, the lectures themselves are often enhanced through the use of computer based presentations, offering several advantages over "traditional" technologies, such as overhead transparencies and blackboards [cf. Lennon, Maurer 1994]. Since networked hypermedia systems like the WWW do not support the creation of presentation material, specialized authoring tools like Microsoft's Powerpoint are used. However, the presented material needs to be accessible to the students after the lecture and should thus be integrated into the same hypermedia system which stores any other documents related to the lecture. Apart from writing a special viewer for each presentation program that is commonly used - an approach too costly -, two degrees of integration may be distinguished.

1. Launching an external application: The presentation is stored in its native format; in order to access it, a viewer for this data format needs to be available at the client side - like, e.g., the standalone viewer of Powerpoint. The presentation may then be downloaded and the viewer started manually, or, more elegant and easier to use, that viewer has been configured as an external viewer for a hypermedia system like HyperWave.
2. Conversion into a supported data format: The presentation may be converted into some "native" data format that is supported by the hypermedia system. In this case, the Powerpoint slides have to be converted into, e.g., a collection of image files in, say, GIF format. This type of integration is also used by tools that convert animated presentations into Java applets.

The major advantage of the first approach is, that providing material is quite not very costly. On the other hand, special viewers are needed on the student's computer, that is on the client side. While these are often available for free, they are also platform dependent: Powerpoint viewers, e.g., are not available for UNIX systems. Another important problem using external viewers is the fact, that the presentation can make no use of the functionality of the hypermedia system: no links into the hypermedia base or from those documents into some specific point within the presentation may be defined.

With the second solution, the preparation of a slide show accessible from the hypermedia system becomes more costly as long as there are no appropriate tools. However, the advantages more than compensate the additional effort. Hyperlinks may be defined from any arbitrary slide, even from specified points within a slide, to any other hypermedia document within the system, that is any other HyperWave or WWW document, and vice versa. Also, the slides may be viewed on any platform - as long as a web viewer exists supporting the chosen data format.

Because of its obvious advantages, we decided to follow the second approach. A tool has been developed that converts PostScript files with minimal DSC information slide by slide into a sequence of GIF images, linking them by default in a linear fashion. It is thus only necessary to "print" the presentation into a PostScript file. With HyperWave, this file might be stored directly - with slow rendering and at the expense of losing all colour information -, whereas with most other WWW services PostScript is only supported by an external viewer; we would thus be ending up in case 1 above. Our tool then uses the HyperWave Interchange Format to create a collection at a specified location on a HyperWave server - supporting a number of options such as resolution, clipping or orientation. The presentation is then ready to use; additional links to and from other hypermedia documents may be added to any of the slides, since GIF is a fully supported HyperWave format. The same method also works with standard WWW clients, where the use of image maps, or a Java aequivalent, is needed for true hypermedia functionality. We have thus found a method for a seamless and simple integration of presentations created with almost any authoring software into open networked hypermedia systems like HyperWave and the WWW.

References

Lennon, Jennifer, and Hermann Maurer: Lecturing Technology: A Future with Hypermedia. Educational Technology 24 (4), 1994, 5ñ14

Norman, Kent L.: The Electronic Teaching Theater: Interactive Hypermedia and Mental Models of the Classroom. Current Psychology: Research & Reviews 9 (2), 1990, 141ñ161.

Development and Evaluation of World Wide Web Courseware for Teacher Education

Isabel Borrás
San Diego State University
5500 Campanile Drive
San Diego, CA 92182-7703
e-mail: iborras@mail.sdsu.edu

The provision of information in electronic format on the World Wide Web (WWW) is not by itself an effective teaching and learning tool. As Fleming (1993) observes, ìGroup work, problem-based learning and self-assessment are all key features of 'deep' learning but have not received sufficient attention from the designers of computer-based independent learning systems." (p. 321). This demonstration deals with the development and evaluation of a WWW package, *RE 503 Internet for Teachers* (Borrás, 1996), which integrates Fleming's suggested learning features. Created for use on a graduate-level course for inservice teachers, the package includes: 1) course syllabus; 2) information about participants' professional background and teaching philosophy; 3) readings; 4) links to on-line resources; 5) on-line feedback and activities forms; 6) instructor's HTML templates for use by participants in their final projects; 7) tools for site evaluation and final projects evaluation; and 8) participants' final projects.

References

Borrás, I. (1996). *RE 503 - Internet for Teachers*. [WWW document]. URL: http://www.sover.net/~mi1casa/503/

Fleming, D. (1993). Opinion: Paradigm lost? *Educational Technology and Training International, 29* (4), 321-324.

Science Technology Entry Program: Integrating Technology into a Cross Curricula Design

Dr. Joseph Bowman, Jr., Assistant Professor
Department of Educational Theory and Practice
University at Albany, SUNY
Center for Urban Youth and Technology
ED 114B, School of Education
1400 Washington Avenue
Albany, New York 12222
518-442-4987
jb219@columbia.edu

Aliya Holmes, Graduate Assistant,
Department of Educational Theory and Practice
University at Albany, SUNY
Center for Urban Youth and Technology
ED 124 School of Education
1400 Washington Avenue
Albany, New York 12222

Donovan Smith, Graduate Assistant,
Department of Educational Theory and Practice
University at Albany, SUNY
Center for Urban Youth and Technology
ED 124 School of Education
1400 Washington Avenue
Albany, New York 12222

Marc Callendar, Teaching Assistant,
University at Albany, SUNY
Center for Urban Youth and Technology
ED 124 School of Education
1400 Washington Avenue
Albany, New York 12222

Etwin Mapp-Bowman, Technology Coordinator
University at Albany, SUNY
Center for Urban Youth and Technology
ED 124 School of Education
1400 Washington Avenue
Albany, New York 12222

Abstract: The Science Technology Entry Program provides math and science for inner city youth around New York State. STEP in Albany, New York is a collaboration between the University at Albany, SUNY, Albany Medical College, and Philip Livingston Middle School. Students meet during the week to learn about math, science genetics, graphing calculators, multimedia technology, and the Internet. Students meet guest speakers and attend field trips in the related fields. The students are developing newsletters, multimedia products, and creating World Wide Web site about the program. The program is exploring ways to integrate technology into cross curricula design areas of math, science, English, technology, and social studies.

Introduction

The Science Technology program at the University at Albany, SUNY has been providing basic skills development for students in middle and high school for six years. The program has used the resources of the university to give these students experiences that they might not have an opportunity to do in their normal school activities. For several years Center for Urban Youth and Technology (CUYT) has developed a body of work that investigates the design, implementations, and research issues of integrating technology into the school curricula. We have focus on the K-12 population, specifically the middle school and high school age range. We believe that has technology is placed in educational environments, it will become part of the reform and/or restructuring efforts that are ongoing in education. It is another opportunity to provide instructional approaches that are novel, exciting and keep pace with the technological capabilities of this time.

Much attention is given to development of math and science skills expertise, but there are other factors that make this skills development possible. Students need to able to read and write in order to understand the math and science concepts that will be presented to them in their classes. Student must have keyboarding, media literacy, and technology skills to access, manipulate, and present data/information in school and in the workforce.

We have initiated several partnerships with the Albany Tri-city area and with New York City schools. In this collaborative environment we have developed several afterschool, weekend, staff development, and parent/community technology-based initiatives. These programs have been designed to provide a test bed for specific questions that CUYT has developed as it relates to multiple curricula design, technology integration into the curriculum, design of technology in school reform concerns with shared decision making teams (SFMT), and multi-variant analysis of student learning in an interactive and student-centered learning environment.

ID Comm: Building a Virtual Community for Instructional Designers

John Dietz
University of Washington

ID Comm is a Web based project designed to encourage collaboration among instructional designers with topics including instructional design, computer based learning, distance learning, Web based learning, etc.

ID Comm includes a database of resources for instructional designers, yet attempts to be more than just another index of resources. ID Comm resources can be accessed in a variety of ways in order to retrieve only the most relevant information. ID Comm also includes a number of forums for discussions on relevant topics.

The success of ID Comm is based upon participation by instructional design professionals. Anyone can easily add resources to the database or add comments to existing resources.

This poster covers the functionality of the ID Comm web site in addition to the design process and the technology involved.

The Remote Item Development Environment (RIDE™)

David Kuntz
Computer-Based Assessment Innovation Team
Educational Testing Service
Princeton, NJ 08541 USA
dkuntz@ets.org

Ned Walthall
Computer-Based Assessment Innovation Team
Educational Testing Service
Princeton, NJ 08541 USA
nwalthall@ets.org

RIDE™ is a prototype of a web-server-based assessment creation tool, developed at ETS. With its three modules, users can easily create computer-based testlets and instructional or diagnostic tutorials, developed through and delivered over the World Wide Web. The **ItemBuilder** allows one to construct pre-formatted test questions, without requiring the user to know HTML. The **Item Registrar** allows one to inform a remote WWW server of the important characteristics of a question, including the correct answer, any explanations associated with the question, supplementary files (links the questions may contain to other sources), and other pertinent information. The **Testlet Creator** allows one to construct linked, scorable collections of registered documents, standard format or hand-crafted, from one's ItemRegistry. Depending on the options selected during the linking process, the testlet will provide score reports, rationales, links to other informational or instructional material, and primitive diagnostic information and branching based on rubrics the user provides.

Preservice Teacher Education and the Web: Shifting Paradigms, New Partnerships

Dr. Deborah A. McAllister
University of Tennessee at Chattanooga
Teacher Preparation Academy
615 McCallie Ave.
Chattanooga, TN 37403 USA
Email: dmcallis@utcvm.utc.edu

Wendy K. Bartnik
Chattanooga Phoenix School II
Science
1219 West Mississippi Ave.
Chattanooga, TN 37405 USA
Email: bartnikw@ten-nash.ten.k12.tn.us

Given the availability of varied resources, it follows that a multidisciplinary/multimedia approach to preservice teacher education is evolving. Methods and materials that incorporate the use of the Internet into the teaching and learning process, at all levels of instruction, are being explored. Student work is accomplished through the use and integration of packaged multimedia products, material located on the web, and content-based print and video material.

With a changing paradigm for preservice teacher education, the higher education faculty role must also change. Partnerships and collaboratives, both formal and informal, have become commonplace in education, and include a wide range of agencies. Unique collaborative training formats are developed for inservice and preservice teachers, faculty professional development, and school-aged children. In Tennessee, a series of technology competencies are met, hence a shifting paradigm of technology, and a new frontier for all partners in the teaching and learning process.

Evaluation of multimedia authoring tool TSUMIKI through developing AIDS courseware

Akiteru Motoda*, Wakako Tokumaru *, Hideo Kiyohara**,
Matsumoto Tsutomu***

* Sensyu University Tamana High School
** Kozo Keikaku Engineering Inc. Kumamoto Offiece
*** Kumamoto National College of Technology

e-mail : matumoto@ecsrv.ec.knct.ac.jp

We have developed the multimedia authoring tool TSUMIKI which targets ordinary K12 educators who are not familiar with computer technology. The TSUMIKI supports them to create courseware for their class and is also useful for self-expression of K12 school kids. Title can easily be converted to Web Page by choosing "HTML" from pull down menu.

In order to evaluate the TSUMIKI, we have developed various coursewares. Stop Aids is one of them. In this paper we try to report evaluation through describing development of courseware STOP AIDS and practical usage of class . The developed courseware consists of twenty seven TSUMIKI pages, fifty narration files and fifty pictures file.

It took very short time to create like this big courseware that we can say TSUMIKI has good user interface to computer novice user.

Un Meurtre a Cinet (Un homicidio en Toluca) *: An E-mail Whodunnit to Develop Writing Competence in Intermediate Level Language Classes Offered at a Distance*

Terri J. Nelson
Department of Foreign Languages and Literatures
California State University
San Bernardino, USA
tnelson@wiley.csusb.edu

Walter Carl Oliver
Department of Foreign Languages and Literatures
California State University
San Bernardino, USA
woliver@wiley.csusb.edu

Un Meurtre a Cinet (Un homicidio en Toluca) is an intermediate level language role-playing project in which students collaborate to solve a murder mystery using e-mail, a listserver, and a Home Page. It is designed to provide students studying language at a distance with an opportunity to work collaboratively in a linguistically and culturally rich context without ever having to meet face to face. The use of e-mail focuses attention on the subtleties of using writing as a communicative tool since students' ability to solve the murder mystery depends upon their being able to ask challenging questions to gain the information they need while avoiding giving answers that inculpate themselves. The project makes extensive use of realia. The use of a home page allows for the distribution of this realia electronically and the use of a listserver facilitates extensive asynchronous interactivity among the students regardless of location.

A Model-in-Progress for Designing Web-based Collaborative Assignments in Business Management Education

James Paradiso
Department of Business Management
College of Lake County, USA
paradiso@clc.cc.il.us

Bipolar forces shape a growing dilemma in higher education. Administrative directives, student surveys, and the media's gold rush mentality about the Internet directly or indirectly power the design of Web-based collaborative assignments for college students. In contrast, the lack of models for designing appropriate learning tasks for information-rich environments and the minimal integration of the Web into teaching and learning are well documented, particularly in the field of business management education.

Attempting to address this demand gap, a preliminary model for designing Web-based collaborative assignments for undergraduate business management majors was developed and beta tested. The model's structure, infrastructural requirements, areas for improvement, and remaining tasks for designers of Web-based collaborative assignments are discussed in a short paper, available on request.

The Human Brain: A Frames-based Image Orientated Quiz

Joan E. Robertson, Ph.D.
Science Instruction Designer
IAIMS
Mail Box # 357155
University of Washington,
Seattle, WA 98195
USA
Email: joan@u.washington.edu

John W. Sundsen, Ph.D.
Associate Professor
Biological Structure
Mail Box # 357420
University of Washington,
Seattle, WA 98195
USA
Email: jws@u.washington.edu

John R Bolles
Associate Director
Health Science Center for Educational Research
T-252 Health Sciences
Mail Box # 357161
University of Washington,
Seattle, WA 98195
USA
Email: bolles@u.washington.edu

Internet resources in medical education are rapidly developing and the Internet can now play a vital role in providing a framework in which medical students can construct their learning collaboratively and actively. This poster suggests that learning can be enhanced further by the use of the Frames feature now supported by most browsers. Within a Frames environment graphics, indexes and forms can become more functional and different types of information can be displayed side-by-side.

In this study 10 quiz modules were designed using a Frames layout. The quiz modules are a series of self evaluation tests on the human nervous system. Feedback from students is discussed in the poster, along with the apparent educational advantages of Frames and Frames based quizzes. The issue of download time versus image quality will also be addressed. Images from the Digital Anatomist Program and further information can be found at: http://www1.biostr.washington.edu/DigitalAnatomist.html .

A new approach in public health education and risk awareness of diabetes at URL http://www.uni-leipzig.de/~diabetes

Hardy Schulze
University of Leipzig, Department of Internal Medicine
Germany
E-Mail: shuh@server3.medizin.uni-leipzig.de

Stefan Bornstein, MD
University of Leipzig, Department of Internal Medicine
Germany
E-Mail: shew@server3.medizin.uni-leipzig.de

Thomas Baehring, PhD
University of Leipzig, Department of Internal Medicine
Germany
E-Mail: baet@server3.medizin.uni-leipzig.de

Diabetes is an important clinical and public health problem. Worldwide, at least 20% and in many communities more then 50% of diabetes is undiagnosed. Therefore a new and simple electronic questionnaire is suited to identify people at increased risk for undiagnosed diabetes type 2 and was designed as a WWW-application. Our CGI-script based test provides a direct feedback in form of personal risk evaluation and a statistics on number of users, on distribution of sex and risk profiles.

The World Wide Web as a standardized communication platform offers an ideal base for a new approach in health education. In combination with innovative techniques this application allows to collect epidemiological data, perform an online disease screening and raise the risk awareness as a major step in health education.

Collaborative Learning/Teaching on the Web: Lessons from Project Vision, Penn State University, 1995-1996

Nancy Wyatt
Speech Communication
Delaware County Campus
Penn State University
United States of America
njw@psu.edu

Susan Ware
Library
Delaware County Campus
Penn State University
United States of America
saw4@psu.edu

Matthew Bodek
Library
Delaware County Campus
Penn State University
United States of America
mjb14@psu.edu

This poster/demonstration session identifies and describes three skill sets necessary for instructors and students to participate effectively in a collaborative, electronic learning situations: communication, information acquisition and evaluation, and computer and Internet technology skills. Faculty and students must learn to collaborate to set learning goals, develop strategies to achieve those learning goals, and respect diverse viewpoints.

Communication Skills

Communication skills include listening to understand, listening to evaluate, speaking and writing clearly, supporting opinions with relevant evidence, asking pertinent questions, responding to and critiquing messages from others, and summarizing discussion.

Information Acquisition and Evaluation Skills

Information acquisition and evaluation skills include identifying information needs, finding information from a variety of sources, and evaluating information for relevance and validity.

Computer And Internet Technology Skills

Computer and Internet technology skills include the ability to access information sources, to use groupware for discussion for communication and discussion, and to participate in video conferences.

Acknowledgments

Information for this presentation was derived from the experience of all of the students, faculty, and staff in the Commonwealth Education System at Penn State University who participated in 1995-1996 Project Vision.

Integrating Servers for Communication Over the Internet

Marco Aiello
Corso di Laurea in Ingegneria Informatica
Universita' di Roma La Sapienza, Italy
aiellom@dis.uniroma1.it

Martin Haeberli
Advanced Technology Group
Apple Computer Inc., USA
(Now at Netscape Communication Corporation, USA)
martinh@netscape.com

Abstract: Communication among groups that share a common interest or goal is a crucial task for all organizations and scientific communities. In many cases, it is also very important that information reaches a certain target audience. We have built a demonstration prototype that integrates different Internet technologies to support communication among groups. We have used the Apple Macintosh environment and glued all the applications together using UserLand's Frontier system-level scripting language [Hale, 1995]. In this article we discuss why it is critical that the same information can be reached in different ways, then we explain our approach to the problem, its strength and its limits.

TV News vs. Newspapers (i.e. Mailing Lists vs. Bulletin Board Systems)

Until now ways of communication on the Internet have included mailing lists, BBS (bulletin board systems), and newsgroups. But which is the best? It is our belief that none and all of them are. What we mean is that there is no winner: each approach has its strengths and its weaknesses. And each user has his preferences.

Some people like to watch the news on TV; some use the radio for this purpose; some like newspapers and others use on-line services; in the same way some people subscribe to mailing lists, some to newsgroups and others frequently search BBS systems. But there is no absolute: if you are used to reading a newspaper every day this does not mean that you are never going to watch the news on TV. If you are in a hurry or you are doing something (like driving) you will probably use the radio; if you want a close look on a topic or you want to see something on a historical event you will read an article on the newspaper (if you still have it).

Up to now administrators just decided which was the most appropriate for the expected needs and end users did not have many choices. It is our belief that the information should be available to you however you wish. Once a discussion is started on the Internet, you must be able to retrieve the contents and to add to them, no matter how you decide to browse and add. If you want fresh information, let email reach you. If you want to search in the line of a discussion, use a BBS that holds all of the notes. If you like web browsers then use your favorite one to reach a site with the messages posted.

Our Proposal

With our prototype, we took a significant step in this direction. We decided to use Apple's Macintosh™ as server. For this purpose we used WebSTAR™ as web server, Apple Internet Mail Server™ (AIMS) for serving email and Macjordomo™ for the mailing lists. Actually, it is possible to use any other mail server on any machine as long as the accounts needed to run our system are created. It is also possible to use other mailing list servers. (For example, the email can be held by one or more UNIX servers running

Majordomo™, with an additional email account for our system). In addition, Eudora™ was used to let the system look for email from the mailing lists and send out the ones corresponding to messages posted on the BBS.

This group of servers was glued together using UserLand Frontier™, a system level scripting language for the Mac. We used an existing web-based BBS from Dave Winer that runs thanks to Frontier's CGIs (Common Gateway Interfaces). The advantage of this type of BBS is in the fancy outline for the messages, the ability to use HTML (HyperText Markup Language) in the body of the notes; since all of this is on the web: this means a larger audience and, often, more friendly interfaces for the readers. An interesting feature of Frontier is its object database (OD) that is always resident in RAM when the Frontier's script is running; this can mean speed and, in many cases, it also means relatively small data bases.

The architecture of our system is as follows: directly on the network we have, on one side, WebSTAR™, which serves WWW requests (HTTP), on the other AIMS which has the duty of moving email to the final receivers and sending outgoing ones.

WebSTAR is invoked when browsing the BBS; each page that it displays is obtained thanks to Frontiers' querying the internal message data base and preparing the HTML pages on the fly. In these pages there are also the forms to fill-in to add new elements to the discussions. Once the forms are submitted, Frontier processes them to update the OD data base. Concurrently with the update, an email is prepared with the same contents of the message posted and is queued (Eudora™ is launched, if not already active, for this purpose). The receiver of the message is set according to the value found in another small database in the OD. The database that is searched is kept with the correspondences between the discussions on the BBS and the existing mailing lists that serve corresponding topics.

On the other side we have the "email part" of the servers: AIMS deals at low level with the email; it is the post office of the system. Macjordomo uses it for reading incoming mail with new messages for the mailing lists. Each message that comes is then copied and sent to each recipient. One of the recipients must be the BBS itself, for which an account is created in AIMS. To check for mail addressed to "BBS", we have built a Frontier™ agent [Maes 1995] that invokes Eudora™ at regular intervals (posting messages on the BBS, creating new mailing lists and moving emails between mailboxes as needed).

We have tested the system by letting it run for almost a week on three existing mailing lists outside our domain and creating two BBS discussions and corresponding mailing list on our server. The system behaved as expected: no messages had any problem when retrieved from the mail agent, no message was lost and all the necessary duplication of information took place (BBS and email version of all messages).

Behind the Corner

This prototype was built for Apple's (Apple Computer, Inc.) internal use. In a company where there are an incredible amount of meetings, events and deadlines, it is likely that managers, researchers, developers or engineers will miss something; but the last thing that the management wants is that people are not productive as much as possible (meaning duplication of work. non-collaboration, non-use of resources due to a lack of knowledge, etc.); so if this happens through improper information circulation (this includes retrieval of obsolete messages or the inability to find old messages), then the problem needs to be solved and solved efficiently.

What we have tried to demonstrate is that there is no standard for communicating. Of course this is only a starting point and there is a lot of work to do in the direction of letting any information being retrievable in all the possible different ways and being sure that nothing is missed or lost.

A point we also wanted to stress is that a working prototype could be built in a relative small period of time thanks to the use of Apple™ technologies. Probably ours is not the most efficient and robust solution to the problem, but certainly it is one which was possible to implement very quickly and which could let us concentrate on our objective, exploring our views on new ways of dealing with the rich styles of communication available over the Internet.

What the authors dream of is a perfect wired world in which, when you turn on your computer in the morning, you can read exactly what you want to read and what you were expected to. Not one single word more. Not one less.

References

[Hale, 1995] Hale, M. (1995). Scripting the Web with Frontier. MacTech Magazine, 11 (12), 57-64

[Maes 1995] Maes, P (1995). Intelligent Software. Scientific American, 273 (3), 84-86

Acknowledgments

The research that lead to this paper was carried on while the first author was an intern at Apple Computer, Inc..

Apple, and in particular Jim Spohrer, are acknowledged for the opportunity.

Converting a Large Introductory Course into an Internet Course: Process, Problems, and Solutions

Gregory F. Aloia, Ph.D. Professor
Department of Specialized Educational Development
Illinois State University
United States of America
E-mail gfaloia@rs6000.cmp.ilstu.edu

Rodney P. Riegle, Ph.D. Professor
Department of Educational Administration and Foundations
Illinois State University
United States of America
E-mail rpriegle@rs6000.cmp.ilstu.edu

Perry L. Schoon, Doctoral Student
Department of Educational Administration and Foundations
Illinois State University
United States of America
E-mail plschoon@rs6000.cmp.ilstu.edu

"We have learned that you have to encourage learning by doing, that the learning has to be relevant to classroom experience ands that you need to model good practice." [Harrington-Luker , 1996].

Converting a college course into an internet course is a sufficiently challenging undertaking in itself. However, when attempting to convert a large introductory lecture section (n>200) into an internet course, there are several additional areas that need to be addressed if the course is to be successful. The course in question is an introductory course for education majors at a large midwestern university known for its excellence in teacher education [Hurst, 1994]. It is the largest course to date in the College of Education to be converted to the internet. The course is an attempt to provide education majors with direct experiences with technology as they prepare for the profession of teaching [O'Neil, 1995].

There are five key areas that must be addressed in preparing to convert a traditional large introductory lecture section into an internet course. They are:

1) The staff and personnel requirements

2) The logistics and support network on campus

3) Evaluation of the course, students and instructor

4) Issues related to individualizing the learning

5) Research questions to be examined

The staff needed for the course would consist of the professor and at least two graduate teaching Assistants (TA's). In addition to the content knowledge related to the course, the TA's will need to be familiar and comfortable with the internet and the technology involved. Specific training will need to be planned to enable the TA's to a) run chat rooms, b) evaluate assignments, c) assist students enrolled in the course with various problems that arise, d) assess new sites on the Web, etc.

Additional contact time for planning and problem solving would need to scheduled between the TA's and the instructor.

The logistics and support network on campus is essential. The demands of 200 plus students needing to not only complete eight to ten assignments but also obtain all of the course materials via the university's computer system can create a serious access problem. Timely and efficient access for the students is essential. Scheduling the labs, having sufficient numbers of lab technicians available, anticipating "system" glitches and making sure that the administrative units (i.e. department, college and provost office) are informed, supportive and willing to provide the network needed to make the course successful.

Negotiating with the chairperson for appropriate release time to complete the undertaking is essential. A conservative estimate is that 250 additional hours will be needed just to initiate the course. Depending on the departmental culture, half to full time assignment is to be sought.

The record keeping and paper transactions for large classes is very time consuming and laborious. A WWW course should relieve some of the instructional burden by providing automated grading and data-gathering through forms and cgi scripts. Securing assistance in this area is essential.

The ongoing and summative evaluation of the course by the students is essential. Since the internet experience may be the first internet course that students have taken in their academic careers, it is important that they are given ample opportunity for providing feedback both during its development and also during its status as a fully interneted course.

As each phase of the course is being placed on the internet the students should be queried for specific feedback on every assignment, both individual and group, as well as the materials (i.e. lectures, references etc.) Areas to be assessed include: Knowledge gained, level of difficulty, usefulness of information, problems encountered, clarity of directions, Ways to improve, Strengths, etc.

In addition to the specific feedback there needs to be opportunities for formative feedback (weekly or biweekly) via e-mail. This would be sought from the students as they read the Web sites, articles, and chapters and complete the individual and group assignments.

A third evaluation or summative feedback would consist of the students assessing the entire course at its conclusion. The current Department teaching evaluation form would probably need to be revised to reflect the nature of an internet course.

One of the traditional concerns about large lecture classes expressed by students and faculty alike is that they are so impersonal. This is especially the case with education and in particular special education. It is an ironic contradiction that the very cornerstone of special education is the individualization of learning for students and we begin the professional training of future teachers with a massive, potentially impersonal learning experience. Through the use of the internet, individualizing the course to reflect the diversity of students and their needs is possible. From the asynchronous nature of the internet, the constraints of both time and location have been addressed. Through the use of chat rooms and e-mail, ongoing, individual and group discussions can occur.

Since the internet is "blind," potential prejudices associated with ethnicity and disability should disappear or at least be minimized and not interfere with the learning. Through the use of test banks students can be tested until a predetermined criterion has been achieved. The pace of the learning can be individualized to account for the many competing demands on the lifestyles of students, many of whom are nontraditional.

Individual mentoring experiences with professional teacher in the field are possible through the internet and e-mail. These experiences would enable students to link up with practitioners and experts in their area of interest for information, advice and direction about the profession of teaching.

In addition to the staff and personnel requirements, the logistics and support network on campus, the evaluation of the course, students and instructor, and issues related to individualizing the learning, there are critical research questions that need to be examined such as: How effective is the internet as an instructional model when compared to the traditional large class instruction versus a hybrid that combines the internet with traditional in-class instruction? Does the potential for individualization of learning and testing improve the learning experience of the student? Will the internet minimize potential biases that historically have existed between teacher and student?

The process of converting a large course to the internet presents many challenges and opportunities for the student and professor. If future teachers are going to enter the field technologically literate professors "need to model good practice."

Literature References

[Harrington-Luker, 1996]. Harrington-Luker, D., (1996) Coming to grips with staff development. Electric Learning, 16 (1) 32-43.

[Hurst, 1994]. Hurst, D., (1994) Teaching technology to teachers . Educational Leadership, 51 (7), 74-76.

[O'Neil, 1995]. O'Neil, J., (1995) Teaching and technology: Potential and pitfalls, Educational Leadership, 53 (2), 10-11.

Bringing Instructional Design Principles On-Line: An Instructional Technologist's Approach to Web Development

Shannon A. Alpert
AT&T, Interactive Products and Services
USA
salpert@attmail.com

The process of Web development employs many instructional design principles. Instructional design principles are apparent in a variety of Web sites including sites that offer information posting, interactive simulations, and training materials. The traditional instructional design model includes analysis, design, development, implementation, and evaluation. This model is often applied to Web design without a conscious awareness of the model. The various parts of this model occur throughout the design and development of a Web site. Knowledge of this model can improve the design and functionality of any type of Web site.

Analysis, the first phase in the instructional design model, is crucial to the success of a Web site. In analysis, Web designers must determine the objective, audience, navigation, features, and advertising strategy of the site. During this phase, plans for each of the subsequent phases are made. Although this may seem like an academic approach to Web development, these determinations are usually made informally. Web designers must have this solid foundation of information from which to work.

The second phase of the instructional design model is design. In this phase, Web designers determine the interface design, links, navigation, and overall structure of the site. Storyboarding is a large part of this phase. On a storyboard, designers can communicate every aspect of the design to the HTML programmers and graphic artists. It is important to note that Web designers guide the design and development of Web sites. However, HTML programmers and graphic artists can provide invaluable input into the design and development of Web sites. Whenever possible, their suggestions should be incorporated into the design of the site.

The third phase of the instructional design model suggests that the process of development continues for the life of the instruction. This is also true for Web sites. Web site development will continue for the history of the site. During development, Web designers have an advantage over more traditional instructional designers. Web designers can review the work of the HTML programmers and graphic artists as it is being completed. This allows Web designers to offer immediate feedback and suggestions for improvement while the site is being developed. In addition, rapid prototyping of Web sites can occur easily by creating the files that make up that site for review before linking those files to the Web.

Implementation is the fourth phase in the instructional design model. Unlike implementing traditional instruction, implementing a Web site is a relatively quick process that involves posting the site on a server. Web designers must also plan for advertising the site during this phase. Effective advertising ensures that the site will attract an appropriate audience of users.

The fifth and final phase in the instructional design model is evaluation. The evaluation process of a Web site is similar to the evaluation process of instructional materials. For instance, instructional multimedia programs and Web sites require a full review of the programming structure and graphic components. Web sites and instructional multimedia programs must also be tested thoroughly before they are released to the public. Ideally, Web sites should be tested using a small group of peers who are observed as they interact with the site. Web designers should also elicit assistance from experts from many fields including Web design and graphical user interface design. Web designers should also select members of the target population to test the site thoroughly.

The instructional design model is a flexible, iterative approach that provides Web designers with the information they need to create effective Web sites. The principles of analysis, design, development, implementation, and evaluation are ongoing throughout the development of the site, and all of the principles should be considered at some level. Applying the instructional design model to Web site design will create a sound, high-quality Web site that will be conducive to updates and engaging to the user.

Acknowledgments

I would like to extend special thanks to Scott Alpert, Skip Atkinson, Vince Eugenio, Marla Nargundkar, Maitland Waters, and Jack Werner.

Watermarking for the HyperWave Hypermedia System

Chris Anderson,
Chris Burns,
Michael Klemme
Hypermedia Unit, Department of Computer Science
University of Auckland
Private Bag 92019, Auckland, New Zealand
chris-a, chris-b, michael-k@hmu.auckland.ac.nz

General

Digital information can be copied without loss of quality, so it is possible to alter documents and incorporate them into new products without leaving a trace. Documents can be transferred over electronic networks with ease and at high speeds. Owners of documents that are of value, which will be made available electronically, need to take precautions to secure their assets. Two basic approaches to prevent misuse of copyrighted material can be identified:

Access control and logging:
Access to the information is restricted to certain users or groups of users. Logging all requests for a document can help in detecting excessive use by a single customer. Unfortunately, this method still does not prevent registered users from passing on documents to others.

Watermarking
Including some additional information in the document allows the provider of the information to prove their ownership and trace the source of ``stolen" information. Watermarking is also known as *labelling* or *fingerprinting*.

With the current exponential growth of the Internet and World Wide Web, much of the information provided (images, text, videos, etc.) may appear on other web servers, without giving credit or payment to the original author. It is obvious that any serious publishing effort on the Internet involving copyrighted material must include some kind of watermarking scheme.

Watermarking Techniques

Visible watermarks change the visible information content of the document and can be seen by the viewer. For images the watermark normally consists of some kind of overlayed image. The pixels of the overlay are computationally merged with those of the original. This process alters the image information and should not be reversible. Removing the watermark from the image will damage it substantially. Watermarks can decrease the clarity and accuracy of an image, but they might be seen as an advertisement of the institution offering the document.

The application of *invisible watermarks* does not visibly alter the document. One of the base technologies used is steganography, which hides additional information in the background noise or randomness of digitized information. This information can not be detected without further information. To be of any use the watermark must still be present, even if the document is modified (i.e. scaled, rotated, cropped, etc.). This can be insured by adding redundant information in the digital signature.

Currently, research groups worldwide are working on sustems using the algorithms described above.

Providing watermarked Images: Our Approach

The ``Watermarking" project was born from the desire for the general public to have access to a large digital image database, whilst safeguarding the images from unauthorised use and duplication. The image database at the time consisted of approximately 1100 full screen images (with associated text data) and thumbnail images. The requirements of the system dictated a scheme whereby certain categories of user could view the unmodified original image, while others view only the watermarked version.

HyperWave provides a mechanism for identification of users and enforces access rights based on this identification on a per-document basis. It provides a container object called a *cluster* which allows selection of the proper version of a document when accessed. This allows for fully transparent use of watermarked or non-watermarked images within the hypermedia system. The image processing part of the system semi-randomly alters and degrades portions of the images' colour and contrast, as dictated by a 1-bit image overlay.

Summary

Whether visible or invisible watermarks are used depends on the requirements of the local installation. Using visible watermarks proved to be sufficient for our application with a public institution like the Museum of New Zealand. It provides enough flexibility for further extension and a growing collection of images. Its access mechanisms make the HyperWave system an ideal basis for storage of both watermarked and non-watermaked images.

References

See our HyperWave Server for further informations and examples.

Michael Klemme
Tue Jul 30 14:48:51 NZST 1996

Impact of the Introduction of Web-Based Teaching on Academic Staff

Pamela Andrew and Graeme Hart
Faculty of Education
The University of Melbourne
Australia

Introduction

Online systems, particularly Internet-based, are making inroads into the more traditional methods of teaching in universities where "new communications technologies such as computer networking enable new approaches to and opportunities for teaching and learning" (Harasim, 1995:K5). The trend to offer courses in distance/off-campus mode is designed to cater for the changing requirements of students, particularly those who find on-campus attendance difficult. While there has been extensive research which examines the needs of the students for distance education, there has been less attention given to the needs of the teaching staff as providers. While needs assessment is an integral part of program planning (Owen, 1993), there are few examples of needs assessment in higher education; this paper is a contribution to the knowledge base in this area.

This paper reports the first phase in a longitudinal study of academic staff in a large faculty of education regarding knowledge of, and attitudes to the introduction of World Wide Web-based online delivery within the Faculty. Strategic objectives adopted by the Faculty for 1997-99 include a goal "to increase the capacity for flexible course delivery (including by distance mode) and to improve the quality of teaching and learning through a rapid expansion in online access to courses using multimedia and in anticipation of widespread access to broadband technologies" (Faculty of Education, 1996: 7). The plan also seeks to "... ensure competence in the use of information technologies of staff ...". A "technology and teaching" committee, with mainly academic representatives from each department, is taking an active role in the program implementing the Faculty's objectives for flexible course delivery, attempting to involve staff in the process of change, avoiding "top-down" and "bottom-up" models of change which may not lead to lasting improvement (Owen et al, 1994).

Initially, all staff were invited to two demonstrations: firstly a demonstration of electronic conferencing as used by the Open University in the United Kingdom, followed by a demonstration of the innovations in online subject delivery of two members of staff. Included in the second demonstration was a presentation of a range of electronic conferencing options.

Following these demonstrations a 36 item pencil and paper questionnaire was completed by staff. This was designed to gauge the knowledge and attitudes of staff before the faculty's program of online delivery began in earnest. The questionnaire investigated four aspects:

* the extent to which staff utilised information technology personally and in their teaching;

* the extent to which staff agreed with statements regarding the potential use of information technology in education;

* possible barriers to the implementation of information technology in teaching;

* perceived needs for training relevant to the introduction of teaching with information/online technology.

Results

In this paper we report on the views of 30 academic staff from across the Faculty about the use of Information Technology (IT).

The greatest current use of IT by academics is the extent to which electronic mail is used to communicate with fellow staff ($\mu = 4.5$) while the use of email to communicate with students is low ($\mu = 1.9$). Since online delivery of educational material is relatively new in the Faculty, the low ($\mu = 1.1$) extent to which students were able to access course material online was not surprising. The low extent to which lectures incorporate use of computer presentation techniques ($\mu = 1.8$) may be explained by the large number of small classes compared with the more traditional large undergraduate lectures and the lack of facilities in some departments. Written responses revealed that some staff were not aware of presentation software and when or where it could be used. Others had not thought of communicating with students via email. This Faculty has provided email accounts (without charge) for all of its students (approximately 4500) as one step in enabling flexible course delivery.

Interestingly, there was support for the notion that IT would both facilitate greater interaction between students and staff ($\mu = 3.6$) and would enhance the teaching/learning experience ($\mu = 3.3$). This was supported by individual comments and was seen by some as essential modelling. In terms of the use of IT in the Faculty, academics were aware that schools are expecting trainee teachers to be literate in the use of IT ($\mu = 4.8$).

Regarding a list of 13 possible barriers to the implementation of IT, time and expertise involved in preparing material for university teaching in new formats rated highest ($\mu = 4.0$) while the perception that IT was a threat to job security rated lowest ($\mu = 2.1$). There was little support for the notion that computer-mediated communications would replace face-to-face teaching ($\mu = 2.8$). Staff expressed concerns about copyright issues and that students without access to computers would be disadvantaged.

The training needs of staff were focused more on the use of bibliographic, electronic mail and web browsers and less on the more traditional database, spreadsheet and word-processing software.

In general, very few respondents questioned the Faculty's strategic plan. The major concern expressed was that online systems should not be just word processed subject notes and communication through email and that, the Faculty should be leaders in this endeavour, and introduce appropriate teaching and learning strategies in the design of course materials. Responses have provided insight into the needs of the academic staff to successfully implement the Faculty's strategic objectives.

Conclusion

The findings revealed that the present use of IT varies enormously across the Faculty. The major use of IT is confined to word processing, and there is also extensive use of email for correspondence between staff. The extent of IT use to this stage has been encouraged by a Faculty wide decision to provide stand alone computers to all staff and support from an active Computer Facility. However, while there have been encouraging developments to this stage, the move to more sophisticated use of IT in teaching, rather than simple communication, is a larger challenge. While the Faculty has developed a Strategic Plan, more work will be needed to implement it. The findings indicate a willingness for at least some staff to begin to use online teaching. However, as Fullan (1987) and others have shown, there will need to be ongoing education and personal level support in addition to facilities in order for extensive online teaching programs to be implemented. Monitoring of these strategies will be undertaken in a further phase of this investigation. As Awbrey has so eloquently put it, "The greatest challenge facing universities ... will be to encompass and encourage the majority of faculty [staff] to move towards the future. Without the majority of faculty, our efforts and broadly integrating new technologies into the curriculum will not succeed, and we will remain preaching to the choir." (Awbrey, 1996:17).

References

Awbrey, S. (1996). Successfully Integrating New Technologies into the Higher Education Curriculum. *Educational Technology Review*, Winter 1996, No. 5.

Faculty of Education. (1996). *Strategic Plan* (draft). The University of Melbourne: Parkville, Vic.

Fullan, M.G. (1987). *Implementing Educational Change: What We Know*. Paper prepared for the World Bank Seminars, Planning for Educational Change. Mimeo.

Harasim, L. (1995). The Virtual University: New Approaches to Higher Education in the 21st Century. In Pearce, J. & Ellis, A. (Eds.) *Learning with Technology*, Proceedings of the Australian Society for Computers in Higher Education, Science Multimedia Teaching Unit, The University of Melbourne.

Owen, J.M. (1993). *Program Evaluation: Forms and Approaches*. Sydney, NSW: Allen and Unwin.

Owen, J.M., Lambert, F.C. & Stringer, W.N (1994). Acquiring Knowledge of Implementation and Change. *Knowledge: Creation, Diffusion, Utilization. 15* (3): 273-284.

Distance Education and the Web in Engineering and Mathematics: Examples and Projects at Harvard, Stanford, and UC Berkeley

Masayasu Aotani, Instructor
Department of Mathematics
University of California at Berkeley, USA
aotani@math.berkeley.edu

Stanford University has been teaching various engineering courses quite successfully over several years while UC Berkeley Extension's EXTEN program and Harvard Extension are also trying to launch a series of courses in engineering and sciences.

Even with today's real time videoconferencing technology and video cameras with fast tracking speed, it is not easy for distant sites to emulate a real on-site classroom. Furthermore, by gathering people at various sites, we are beating one of the purposes of distance education; namely elimination of time spent commuting to the school site. In order to address these shortcomings, universities across the nation are beginning to distribute teaching materials through the web and making parts of lessons available to students over the net so that they can study when they have time and where they are comfortable.

Harvard Extension's Mathematics courses rely heavily on a software called Mathematica, Stanford University's engineering distance education is basically videoconferencing, and UC Berkeley's project aims to incorporate real-time lectures, web, videotape, and snail mail.

Summarized below is a rough sketch of how it is structured at UC Berkeley at this relatively early stage of their project. The target audience is anyone with enough background to take first semester calculus.

1. Advertising
   Snail mail, email, web
2. Application Process
   Both email and snail mail applications will be accepted.
3. Textbook
   ``Real book''; i.e. a hard copy
4. Examples including those not covered in the book
   (updated regularly)
   Available at our website
5. Homework
   Available on the net
6. Homework Solutions
   Available at our site
7. Summary of Lectures
   Summaries of all lectures to be given through the semester are available at the beginning of the semester. This is to allow the students to proceed at their own pace.
8. Lectures (Telecast to various sites)
   This is basically videoconferencing. Attendance is optional.
9. Examinations
   Supervised on-site examinations

Advantages and limitations of teaching a distance mathematics course in this manner are as follows.

The advantages include:

1. No need for daily commuting to school
2. Ease of reviewing
3. Self-paced studying
4. Low cost to students
5. Easier time management

But, the limitations are:

1. Lack of individual attention
2. Very low level of student-to-student interactions
3. Need for reliable local coordinators
4. Limited student-teacher interaction
5. Possible communication problems between the lecturer and the site coordinators
6. Difficulty in maintaining uniformity in the quality of instruction from site to site

Japanese Instruction over the Web: Examples at MIT and UC Berkeley

Masayasu Aotani, Lecturer
Technical Japanese Language Project
Massachusetts Institute of Technology, USA
aotani@mit.edu

Many institutions across the United States are using the internet and the World Wide Web in innovative ways to teach Japanese. Some of the examples of using the net for Japanese instruction are given below.

1. Communication with e-mail pen pals in Japan
   UC Berkeley, Tennoji High School, Wellesley
2. Collaborative publication of school papers
   Wellesley
3. Sharing syllabus over the net
   MIT
4. Development of common resources
   University of Washington
5. Distribution of teaching materials over the net by the National Institute for Educational Research in Japan
   NIER, Tsukuba University
6. Japanese Cafe with live conversation
   Infowave
7. Japanese phonetic drill over the net
   Fujitsu

This trend which started at the college level is now spreading down to secondary and primary education. For example, Professor Miyagawa's group at MIT is organizing local high schools in an effort to promote a wide use of the net for language instruction.

Here in the Bay Area, the Instructional Technology Program at the University of California at Berkeley is helping the Japanese Program in the East Asian Languages Department in the areas of Japanese e-mailing and preparation of Japanese homepages.

With the embedded Japanese capability of Netscape and other browsers, using Japanese on computers and displaying characters on the screen are becoming effortless these days.

It is becoming quite easy and increasingly popular to make the cyberspace a part of the classroom.

At a more advanced level, there is the famous Caters sentence structure analyzer developed by Tsukuba University for MIT's Technical Japanese Language Project.

This software shows the basic skeleton of each Japanese sentence in selected articles along with readings of all kanji. It is intended for use by fourth-year level students of Japanese with technical background. Though it cannot replace teachers, this is a hefty package to be used both in preparing for classes and in reviewing the course material.

Robots And The Web: An Electronic Shopping Agent

Bassam Aoun
IMAGE Technology Research Group
Curtin University of Technology
Perth, Western Australia
Email: saounb@cc.curtin.edu.au

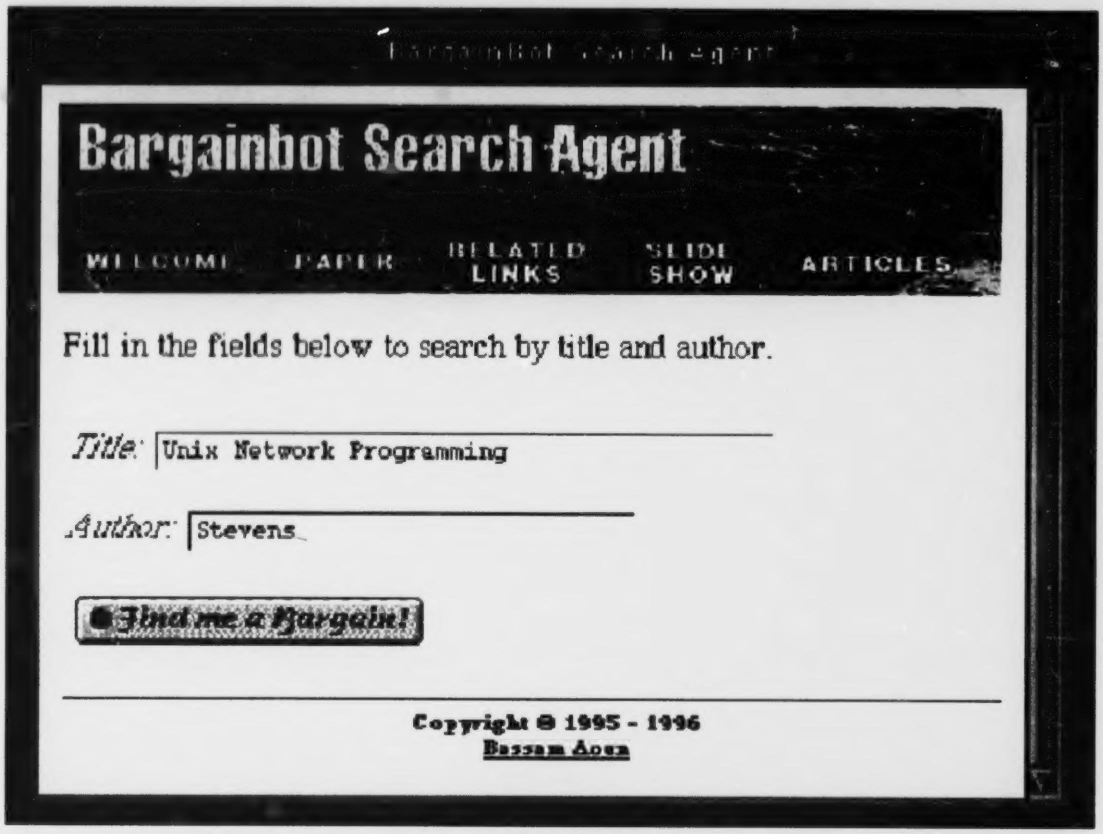
Introduction

Automated, autonomous browsing has an increasingly important task in information discovery and assisted browsing on the Internet. Where users could once keep up to date with information of interest on the Internet, the exponential growth of the network has made this process increasingly time consuming and less rewarding.

Automated browsing can be performed by a number of software entities including agents, robots, infobots, knowbots, wanderers, spiders, crawlers, worms and ants. The application of such programs include indexing, link/html validation, mirroring, electronic commerce, monitoring changes to resources and resource discovery.

Bargainbot Shopping Agent

Bargainbot [URL 1] is a Web-based electronic shopping agent that helps users search for particular product items on the Internet. Bargainbot is currently being trialled with online bookstores. Bargainbot's multi-threaded, multi-connection architecture facilitates simultaneous searching of multiple bookstore databases on the Internet. Bargainbot filters and formats its findings in a manner facilitating comparison shopping by the user. Figure 1 illustrates Bargainbot's Web front-end. Upon specifying the title and/or author, the agent searches multiple bookstore databases and presents its findings to the user. Returned results include the cost of the book and a hypertext link to the appropriate store where the user can purchase the book (Figure 2).


BargainBot Search Agent
Bargainbot Search Agent
WELCOME
PAPER
RELATED LINKS
SLIDE SHOW
ARTICLES
Find me a Bargain!

Fill in the fields below to search by title and author.

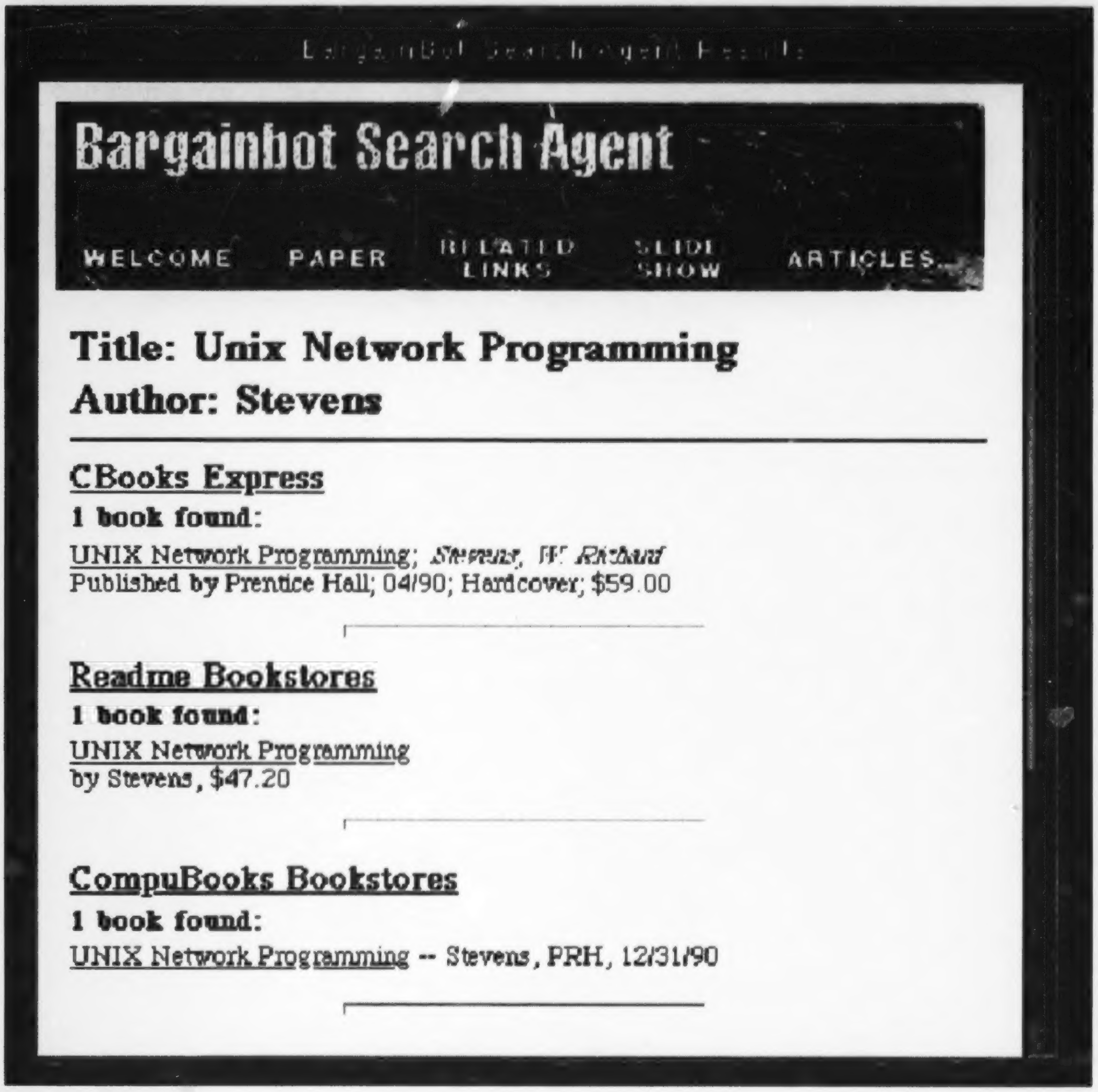
Title: Unix Network Programming

Author: Stevens

Copyright © 1995 - 1996
Bassam Aoun

Figure 1. Screenshot of BargainBot's Web front end

BEST COPY AVAILABLE


BargainBot Search Agent Results
Bargainbot Search Agent
WELCOME
PAPER
RELATED LINKS
SLIDE SHOW
ARTICLES...

Title: Unix Network Programming
Author: Stevens

CBooks Express
1 book found:
UNIX Network Programming; *Stevens, W. Richard*
Published by Prentice Hall; 04/90; Hardcover; $59.00

Readme Bookstores
1 book found:
UNIX Network Programming
by Stevens, $47.20

CompuBooks Bookstores
1 book found:
UNIX Network Programming -- Stevens, PRH, 12/31/90

Figure 2. BargainBot Search Results

To ensure shortest possible response times, Bargainbot is based on a multi-agent architecture, with sub-agents activating individual network connections with bookstore databases worldwide (Figure 3). Individual sub-agents have prior knowledge of how there associated site formats its replies, allowing it to extract the necessary information.

BEST COPY AVAILABLE

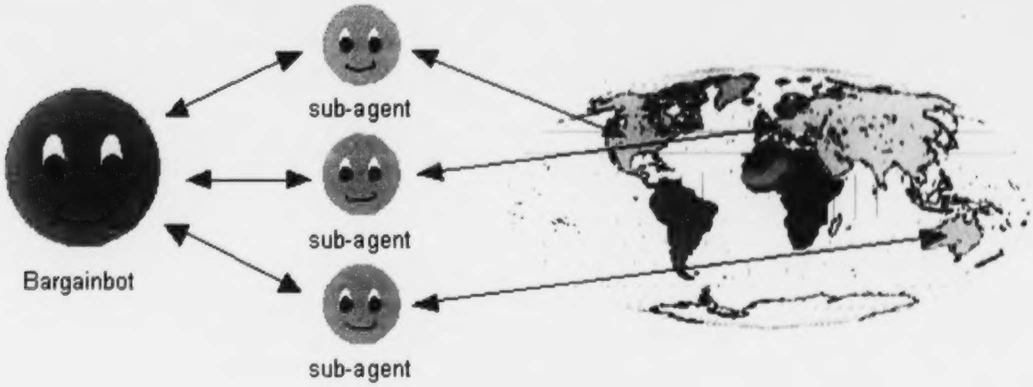


Figure 3. BargainBot parallel, multiagent architecture

Findings and Issues

Purchasing books with the aid of Bargainbot has time and money saving advantages. Performing similar tasks would entail tracking down and searching multiple bookstores on the Internet, a time consuming process.

Eichmann [Eichmann 1994] examines the ethics of Web-based agents. Bargainbot does not suffer the inefficiencies of recursive retrieval associated with conventional robots. Bargainbot knows exactly where to locate its information and does not produce recursive hits on a server or excessive stress on the network.

The development and use of Bargainbot has highlighted numerous issues that need to be addressed if automated browsing, and specifically automated querying of site databases, is to remain a useful complement to manual browsing on the network.

Many online stores provide little support for effective information querying of their databases. For example, many bookstores that Bargainbot searches, have a Web forms front-end to their database that limits querying to either title or author, but not both simultaneously. This requires Bargainbot to perform an extra level of filtering to remove unwanted authors from a returned list of book titles. Further filtering is required to remove redundant markup (HTML) from returned responses. The format of data returned from bookstore sites also varies. Bargainbot must apply different filtering mechanisms to different sites in order to return combined results in a homogeneous manner. Furthermore, changes to the format of bookstore responses requires programmer intervention. The use of machine learning techniques to sense and adapt to such changes will be crucial to the success of similar agents in future filtering tasks.

Other problems associated with automated browsing of the Web include the presence of client-side state mechanisms to store state information on the client side. State information, whilst useful for manual browsing of interconnected pages, is often meaningless in the context of automated browsing, resulting in agents having interface problems with site databases that expect such information. Certain sites also make use of session keys, which time-out after short time intervals, requiring their continual renewal.

The above problems could be overcome by the introduction of a suitable merchandise request protocol. Other possible querying mechanisms include Z39.50 [Kunze] or an inter-operable agent model [Genesereth et al. 1994]. Technical cooperation amongst site administrators and agent maintainers is also needed for the effective use of agent technology in electronic commerce. For example, certain sites blocked the BargainFinder agent [URL 2] from searching their databases.

The use of Java [Gosling & McGilton] as an agent implementation language is being looked at by many researchers [URL 3] [URL 4]. Client side execution of agents have numerous advantages. For example, client-side execution of network operations overcomes traffic routing that server based filtering mechanisms produce, hence placing less stress on the network. Figure 4 illustrates Bargainbot interrogating remote bookstore databases, filtering its findings, and returning the information to the user's browser. Figure 5 shows the network operations of a client side agent, such as one written as a Java applet. Note there is only one set of network operations required.

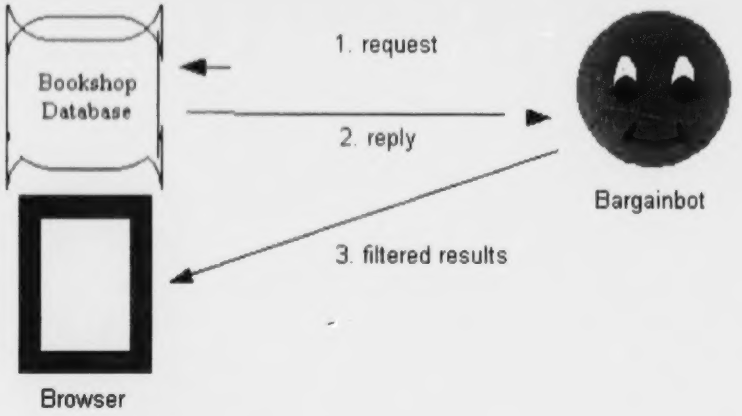


Figure 4. BargainBot network operations

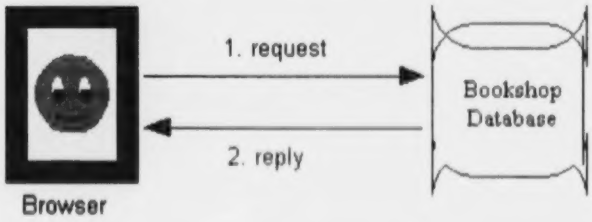


Figure 5. Client-side agent network operations

Conclusions and Future Work

Bargainbot is undergoing continual refinements including enhancements to its responses, further facilitating comparison shopping amongst different stores. A Java rewrite is also in progress.

Electronic shopping agents like Bargainbot have a pertinent place in electronic commerce. However, before we can expect the widespread use of such agents, various issues need to be addressed. For example, Bargainbot and BargainFinder operate well for products such as CDs and books. However, future electronic commerce agents should also consider the importance of more subjective criteria such as product quality, after sales service, warranties etc. Furthermore, future agents should be intelligent enough to seek new avenues for information and adjust to new environments. The personalisation of agent technology should also be investigated. An agent that learns of the shopping habits and interests of its user, and adjusts its operations accordingly is of more worth.

References

[URL 1] - http://www.ece.curtin.edu.au/~saounb/bargainbot/

[URL 2] - http://bf.cstar.ac.com/bf/

[URL 3] - http://www.mcc.com/projects/infosleuth/papers/intranet-java.html

[URL 4] - http://cdr.stanford.edu/ABE/JavaAgent.html

[Eichmann 1994] Eichmann, D. (1994) Ethical Web Agents, Proceedings of the 2nd International WWW Conference, Chicago, IL, October 1994.

[Kunze] Kunze, J. A. (1992) Nonbibliographic Applications of Z39.50. The Public-Access Computer Systems Review 3(5), 4-30.

[Genesereth et al. 1994] Genesereth, M. et a. (1994) A Distributed and Anonymous Knowledge Sharing Approach to Software Interoperation", Computer Science Department, Stanford University

[Gosling & McGilton] Gosling, J. & McGilton, H. (1995) The Java™ Language Environment: A White Paper, Sun Microsystems

Acknowledgments

Financial support and assistance from the IMAGE Technology Research Group & Curtin University of Technology is gratefully acknowledged.

Active Messaging in Java

Alireza Bahreman
Project Leader, EIT/VeriFone
<bahreman@eit.com>

Nick Zhang
Senior Software Engineer, EIT/VeriFone
<zhang@eit.com>

Introduction

Currently, web browsers retrieve data and documents mostly in a "pull" mode. That is, information is presented to the user upon accessing a particular site. On the other hand, email is mostly used in a "push" mode in which information is sent to the user, not necessarily involving their initiative. The ability to extend the information access using the push model expands the application and utility of the Information Infrastructure.

In addition to static or "passive" data on the Net, it is desirable to be able to retrieve and/or send "active" content. As an example of active content, consider the applets written in Java, which are becoming a household name on the World Wide Web. As a matter of fact, the idea of interactive or active messaging has been proposed and experimented in the email community for some time [ENABLED-MAIL].

Combining the need for active content with the push model for information delivery is exactly what is being advocated here. This paper discusses how to enable mail user agents as well as web browsers to send and display an active message. In particular, we describe how to encapsulate an applet object in MIME. Since the

Web also uses MIME content type to identify document types, this approach is promising in both the push mode for email and pull mode for the Web.

We also discuss how other protocols for adding security to MIME objects could be used to provide additional security. There are two types of security for messaging. They are privacy (encryption) and authentication (signatures). In most cases of active messaging, authentication is the most desired. This is because using a signature the user can verify the identity of senders or authors of the applet.

MIME Encapsulation of Java Applets

We propose a new MIME subtype of the Application Content-Type called, Applet, to encapsulate information needed to transport an applet. In the MIME header of the applet object, there may be several parameters. For example, in order to identify the class names for the applet and related classes embedded in the Application/Applet, a 'name' parameter must appear in the Content-Type field. This will enable applets and related classes to be saved as their real class file names before an applet player is invoked.

Another parameter is the 'site' to which the applet is to connect. This information is needed to be conveyed to the application processing the applet which determines whether or not to grant such privilege. Since not all applets may require to connect to other machines, this parameter is optional.

The simplest example to demonstrate the use of the Application/Applet Content-Type is a single self-contained applet. The parameters specify the original name it was saved under (MyApplet.class), and the site to connect to. The body of the message then contains the base64 encoded applet class file.

Content-Type: application/applet;

name="MyApplet.class";

site="panoply:8080"

Content-Transfer-Encoding: base64

<base64 encoded class file>

In general, applets are more complicated. Some require input parameters and some refer to non standard classes. The Multipart/Related MIME Content-type [RELATED] could be used as a container to include the applet, its input data as well as other referred classes. Since the applet can be considered a data object within an HTML object, we envision a possible tight coupling between the MIME encapsulation of aggregate HTML documents [MHTML] work and the work presented here. For example, just like an image data would be included in a MIME encapsulation of an HTML object with a <IMG> tag, an applet data could be included for an <APPLET> tag. For more detailed information, please read the full version of this paper [MAPPLET]. Also please refer to our web site <http://eco.eit.com/mapplet/> for the latest status of our implementation.

Using MIME Security to Sign Applets

The MIME encapsulation of an applet object is vulnerable to the same attacks as any other MIME object. In particular, it is possible for an active eavesdropper to modify the MIME message in transit. Therefore, neither the authenticity of the sender nor the integrity of the applet can be relied upon. The content of the MIME message may be seen by any passive eavesdropper. To combat these security concerns, the sender of the MIME message could sign and optionally encrypt the MIME message sent. The signature can be verified by the recipient (assuming the public key of the sender is available) assuring the authenticity of the sender and the integrity of the message contents. If encryption is performed, the recipient can decrypt and be assured that no intermediary was able to read the content of the message. Encryption provides for confidentiality of the applet during transmit.

The framework within which security services may be applied to MIME body parts is described in [RFC1847]. One can use Multipart/Signed and Multipart/Encrypted Content-Types to secure MIME objects. As an example, consider the use of the MOSS mechanism. To sign or encrypt the applet, the applet class files are encapsulated in Multipart/Related object which is in turn signed by MOSS and put into a Multipart/Signed object. When the recipient gets the applet, they should verify the MOSS signature first before invoking any application to run the applet. If the signature verification fails, the recipient mail user agent should prompt the user a warning dialog before proceeding to the next step of invoking and/or executing the applet. Other security considerations should follow the Java applet security guidelines. For example, the Java applet and related classes should not be saved into a directory which is specified in the CLASSPATH environment variable.

Comparison with Other Approaches

In [ENABLED-MAIL], the enabled or active messages are classified into four categories based on four delivery phases: delivery-time, receipt-time, activation-time, and submission-time. The model presented in this document can be regarded as activation-time active messaging, which is similar to Safe-Tcl approach in this sense.

Compared to Java, Safe-Tcl lacks general purpose and interoperability with WWW. There are several advantages of choosing Java as a platform for active messaging, mainly 1) inherent security check in Java Applet, 2) platform independence, and 3) uniform graphic interface.

Conclusions

Embedding Java applets into a MIME object can serve both the push and pull modes well. In addition, MIME/Java can take the advantage of several MIME based security mechanisms already in place, for example, MOSS [MOSS]. Therefore, our approach to active messaging using MIME encapsulation is very promising.

Acknowledgments

The authors would like to thank Jim Galvin for his contributions and Rajkumar Narayanaswamy for doing part of the implementation.

References

[ENABLED-MAIL] N. Borenstein & M.T. Rose. "MIME Extensions for Mail-Enabled Applications: application/Safe-Tcl and multipart/enabled-mail", Draft in preparation, First Virtual Holdings, Dover Beach Consulting, September 1993.

[MAPPLET] A. Bahreman, et al. "MIME Encapsulation of Aggregate Applet Objects", Internet Draft <draft-bahreman-mapplet-spec-00.txt>, http://eco.eit.com/mapplet/mapplet.html, June, 1996.

[RELATED] Harald Tveit Alvestrand & Edward Levinson. "The Mime Multipart/Related Content-Type", Internet Draft <draft-ietf-mhtml-related-00.txt>, May 1996.

[MHTML] Jacob Palme & Alexander Hopmann. "Mime E-mail Encapsulation of Aggregate HTML Documents (MHTML)", Internet Draft <draft-ietf-mhtml-spec-00.txt>", April 1996.

[RFC1847] J. Galvin, S. Murphy, S. Crocker, & N. Freed. "Security Multiparts for MIME: Multipart/Signed and Multipart/Encrypted", October 1995.

[MOSS] S. Crocker, N. Freed, J. Galvin, & S. Murphy. "MIME Object Security Services", Internet RFC 1848, October 1995.

Design, Development and Delivery of Instructional Materials over the Internet

Brenda Bannan
Instructional Technology
George Mason University, USA
bbannan@gmu.edu

William D. Milheim
Instructional Systems
Penn State Great Valley, USA
wdm2@psu.edu

Introduction

The World Wide Web is becoming one of the fastest growing resources available to a wide variety of professionals today. Its vast assets are being utilized more and more for research, data collection, communication and other similar tasks in nearly every professional field. In fact, in many ways it is revolutionizing the ways many people carry out their required tasks in a variety of different work environments.

In education, however, the Web is only beginning to be utilized in any substantive manner for instructional purposes. While teachers are beginning to use Web-based resources for research and/or information-gathering, they have not yet begun to use the instructional potential that the Web can offer for actually teaching students in diverse locations or at times different than when a teacher is currently available (asynchronous).

The purpose of this paper is to provide specific strategies for utilizing the Web for teaching rather than simply as a resource for data gathering. These strategies should assist faculty in a variety of educational settings in their use of the Web to provide instruction outside of traditional classroom environments.

Suggestions for the Design of Web-Based Instruction

As with any other instructional products, the design of the materials is one of the most important steps in the overall process. In the case of web-based instruction delivered at potentially great distances from the assigned faculty member, the design is even more critical since the materials themselves must carry much of the instruction with only limited human intervention in the educational process.

Based on these needs, the design of web-based materials should include:

1. a significant audience analysis, since the learners for this instruction may be quite diverse and geographically dispersed,

2. an interface design that allows for effective viewing on the computer screen as well as high quality printed copy,

3. frequent student-to-instruction interaction due to the individualized nature of the this medium and the inherent potential for the use of "page-turning" software, and

4. frequent assessment to provide feedback to the student since actual human contact may be quite limited.

Suggestions for the Development of Web-Based Instruction

The effective development of web-based instructional materials is also critical to the successful implementation of educational experiences provided over the Web. While this development process for traditional instructional materials may include a number of different processes and media formats, it is somewhat limited for Web-based delivery to text, graphics, audio and video presentations.

The development of these materials should include:

1. the utilization of several specific job skills including HTML programming, audio and video production techniques for software applications available for Web-based dissemination, and generic cross-platform computer skills for evaluating the effectiveness of the materials on several different computer systems.

2. comprehensive evaluation strategies for assessing the functionality of the materials in multiple computer environments, primarily Windows and Macintosh, and

3. significant pilot testing of the materials due to the wide variety of computer skills in the potential learners for a given program.

Suggestions for the Delivery of Web-Based Instruction

Finally, the delivery of educational materials over the Web includes a number of factors based on the unique characteristics of this medium. These include:

1. dissemination of the availability of the instruction (e.g., registration with Web-based search engines) for those educational experiences design for more general audiences,

2. the availability for downloading materials that may be helpful to individual learners (e.g., graphic files, outlines, etc.), and

3. placement of the materials on a server that permits high speed access for multiple learners.

Summary

Web-based instruction is gaining increasing attention as a potential medium for the delivery of high quality educational materials to a diverse audience. While this medium does have great potential for the effective presentation of instructional materials, the overall process must include effective design, development, and delivery strategies for the materials delivered in this manner. It is hoped that the above suggestions will help all instructional personnel interested in this new methodology for the delivery of high quality courseware.

ALCOM/NIST Phase Separation Project

Laura Bartolo
Assistant Professor
Libraries & Media Services
NSF Science & Technology Center for Advanced Liquid Crystalline
Optical Materials
Kent State University
Kent, Ohio 44242 USA
LBARTOLO@LMS.KENT.EDU

Outline: Phase Separation Project & Information Technology Component

Scientific enterprise represents a pivotal area in research collaboration because of its prolific use of emerging information technology and its impact on technology transfer. This paper describes the development of a web-based information communication/tracking tool as part of a large multidisciplinary scientific research project involving research scientists, government agencies and industrial partners. The paper will address assessing the impact of the information tool on the scientific research project as well as partnering with research scientists, government agencies, and industrial partners. The anticipated result of the information component of the ALCOM/NIST scientific research project is the development of a new information communication and management tool

The information technology component of the ALCOM/NIST Phase Separation Project will address the following needs of remote scientific collaborations:

- to construct a Web server to contain scientific information
- to direct a user to specific kinds of information,
- to track the generation flow and usage of relevant information, and
- to facilitate collaboration among government researchers, university faculty and industrial partners participating in the project distributed throughout the country.

ALCOM/NIST Phase Separation Project

The Phase Separation: Dynamics and Morphology research project is an interdisciplinary, effort involving physicists chemists, mathematicians, and information specialists from the following institutions: NSF Center for Advanced Liquid Crystalline Optical Materials (ALCOM, Kent State University), IBM, National Institute of Standards and Technology (NIST), New York University, UCLA, University of Toronto, Wright Patterson Airforce Base. The goals of the Phase Separation project are described below:

Phase separation dynamics determines the structure and performance of liquid crystal/polymer materials. Key issues involve polymerization-induced phase separation, and the coupling between orientational order and the emerging morphology. Experimental tools to probe the structure factor and kinetics will be X-ray and light scattering, calorimetry, microscopy to study polymer dispersed and polymer stabilized systems. Theoretical models will be developed and high performance, massively parallel computer simulations will be performed.

Information Dynamics

The information communication/tracking tool will assess the impact of the web server on the Phase Separation research project as well as to provide a data exchange mechanism among research scientists, government agencies, and industrial participants. All electronic information generated during the life of the Phase Separation Project will be collected, organized and archived. An automated forms service will enable project participants to directly contribute to the project database. Users will be directed to relevant information via a search engine and current awareness mechanism. Electronic whiteboard and videoconferencing will support telecollaboration. This work aims to help project participants as well as the scientific community to locate and disseminate project information rapidly in order to facilitate technology transfer. In addition participants will be alerted to relevant new scientific developments outside of the Phase Separation Project in order to improve research for developing new technology.

As large multidisciplinary scientific research projects involving research scientists, government agencies and industrial partners proliferate, tools to improve the efficiency of information flow within collaborative scientific projects and to enhance government and university technology transfer efforts must be constructed. The information technology component of the Phase Separation Project can serve as a pilot for the development of an effective information communication/tracking system.

Deliverables

The web server will develop an information communication/tracking tool to facilitate rapid site-to-site collaboration in finding, creating, and communicating research results among project participants:

Information Communication Mechanism:

To facilitate collaboration among the Phase Separation Project participants, this project will create an database/archive which collects, indexes, abstracts and updates project information generated by researchers, such as the content of scientific presentations, preprints, program codes, experimental data, visualization, modeling and simulation results. Participants will contribute directly original research to the database through web-based forms. Another feature of the database/archive will be an automated SDI (Selective Dissemination of Information) current awareness mechanism via e-mail based upon personalized research profiles submitted by all project participants. In addition, a site-based search engine will enable the database/archive to function as a collective repository for informal information (personal correspondence, notes), formal information (presentations, preprints), program codes, and data produced by researchers in the project. An electronic whiteboard will be incorporated into the Communication/Tracking tool for drawing, writing and videoconferencing applications to approximate face-to-face interactions among users.

Information Flow Tracking Mechanism:

The web server will be constructed to support a distributed, open system framework to share research information, as well as modeling, simulation, and visualization applications running on various platforms. To monitor the flow of information among project participants, unobtrusive tracking mechanisms will be utilized to measure how the users acquire, organize, maintain and retrieve project-related information throughout the life of the Phase Separation Project. Usage tracking of such features as database contributions, SDI current awareness, database/archive search engine , and whiteboard teleconferencing will show how a user was directed to scientific information generated by the project and to document what was selected as relevant information.

Timeline

October 15, 1996 Modify existing ALCOM/NIST Phase Separation Web site to include searching and SDI current awareness services.

April 15, 1997 Implement tracking and monitoring usage mechanism.

October 15, 1997 Provide preliminary assessment of tracking and monitoring usage mechanism.

May 30, 1998 Lay out recommended information flow to improve data exchange and technology transfer efficiency in collaborative scientific research projects.

Acknowledgements:

Financial Support for this project was provided by the National Science Foundation Center for Advanced Liquid Crystalline Optical Materials (ALCOM, Kent State University) and the Center for Theoretical and Computational Materials Science at the National Institute of Standards and Technology (NIST).

Internet for Teachers: Case Study of a Graduate Course

Isabel Borrás
San Diego State University
5500 Campanile Drive
San Diego, CA 92182-7703
e-mail: iborras@mail.sdsu.edu

Introduction

Research suggests that there is a dramatic need for inservice and pre-service professional development opportunities for K-12 teachers in technology (Rice, 1995). Traditional approaches to integrating new curricula into the classroom will not be successful when applied to technology. As Grandgenett and Mortenson emphasize, "Merely supplying teachers with technology often does little good unless the teachers are also carefully trained to use the technology through an appropriate inservice program." (1993, p. 56).

The study outlined here investigated the effectiveness of a graduate-level course for providing teachers with extensive training in the educational uses of Internet technology.

Course Summary

Internet for Teachers was a three-credit, graduate-level course designed and taught by the author at Bennington College. Participants in the course, eight K-12 teachers and three college students, developed proficiency in using various Internet cognitive tools and created six on-line WWW basic projects for use in K-12 education. While learning the cognitive tools, participants developed the outline of their projects and wrote the projects' storyboards. Participants moved then into the development stage: they learned the basics of Hypertext Markup Language (HTML); they put their projects into HTML format, using a number of templates developed by the author; and they field-tested their projects with K-12 students. Participants invested an average of 10 weekly hours of computer work during 14 weeks. The course followed a traditional classroom-based format but was supported by a WWW package, *RE 503 - Internet for Teachers*, developed by the author (Borrás, 1996).

Goals

The study sought to investigate the effectiveness of the course in terms of outcomes and processes. In terms of outcomes, the study aimed at finding out if participants would be able to achieve the main objectives of the course: 1) to learn the use of major Internet tools for communication, information research, and site evaluation; and 2) to author basic educational Web projects for use in K-12 education. In terms of processes, the study intended to investigate the causal relationship, if any, between matching features of the course learning environment and webware materials and participants' achievement of the course objectives.

Theoretical Framework

The study adhered to the tenets of constructivist theory (Bodner, 1986; Perkins, 1991; Jonassen, 1991). According to Jonassen, constructivist learning environments are most appropriate for the second of the three stages in knowledge building (introductory, advanced, and expert). "It is at the advanced knowledge acquisition stage that learners acquire the knowledge required to solve complex, domain- or context-dependent problems." (p. 29). In light of Jonassen's assertion, the adoption of a constructivist framework seemed all the more appropriate since the participants in the study were adult learners. It was expected that, as a result of their many years of experience, teachers in this course would be equipped with the prerequisites for advanced knowlegde acquisition: problem solving skills, interest, autonomy, and cooperation.

Data Sources

Using a case study research methodology (Yin, 1994) data were gathered from the following sources: 1) teachers' reflections on their learning of the Internet tools; 2) students' mid- and final course evaluation summary; 3) students' reports of users' evaluations of the course projects; 4) students' feedback on the course; 5) instructor's evaluation of the course projects; 6) instructor's in- and out-of-class observations; and 7) e-mail messages from course participants and non-participants.

Results

In analysis of the data, information about outcomes and processes of the course emerged. Data revealed the effectiveness of the course to empower teachers to use Internet tools and to author Web projects. In the same way, data shed light about the causal relationship between the implementation of specific learning environment strategies mirrored by specific webware design features, and the course intended outcomes. The five pairs of environment strategies and webware features included: 1) provision of effective guidance/simplicity of structure and ease of navigation; 2) tailoring of information to learners' needs/ originality of content; 3) promotion of intensive, meaningful practice/variety of task-based activities; 4) fostering of in-class working collaborations/diversification of on-line collaborative opportunities; and 5) favoring of learners' participation in design of course materials/ plurality of authoring sources.

Educational Impact

The educational impact of this study is, basically, threefold. First, the study provides a working model for future Internet-related courses for pre- and in-service teachers: if teachers are to integrate telecommunications technologies in their teaching practices, teacher education courses should feature the teaching strategies and the delivery tools that would allow practitioners to "learn by doing." Second, the study promotes the soundness of an important trend in education, participatory design (Silva and Breuleux, 1994) for which attention to the learner and teacher, as opposed to the "expert," is primary. And third, at a moment when limitations in instructional design models are being debated (Hannafin, 1992), the study may help to conceptualize more inclusive and flexible learning environments: it may empower teachers to develop and assess the effectiveness of on-line resources whose design features are patterned on selective teaching and learning strategies.

References

Borrás, I. (1996). *RE 503 Internet for Teachers*. [WWW document]. URL: http://www.sover.net/~mi1casa/503/

Bodner, G. M. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education, 63*(10), 873-878.

Grandgenett, N., & Mortenson, R. (1993). A collaborative inservice model for training teachers in advanced technologies. In D. Carey, R. Carey, D. A. Willis, & J. Willis (Eds.). *Technology and teacher education annual.* (pp. 55-59). Charlottesville, VA: Association for Advancement of Computing in Education.

Hannafin, M. (1992). Emerging technologies, ISD, and learning environments: critical perspectives. *Educational Technology Research and Development, 40*(1), 49-63.

Jonassen, D. H. (1991). Evaluating constructivistic learning. *Educational Technology*, September, 28-33.

Perkins, D. (1991). Technologgy meets constructivism: Do they make a marriage? *Educational Technology, 31*(5), 18-23.

Rice, M. (1995). Issues surrounding the integration of technology into the k-12 classroom: Notes from the field. *Interpersonal Computing and Technology Journal, 3*(1), 67-81.

Silva, M., & Breuleux, A. (1994). The use of participatory design in the implementation of internet-based collaborative learning activities in K-12 classrooms. *Interpersonal Computing and Technology Journal, 3*(5), 99-128.

Yin, R. K. (1994). *Case Study Research, Design and Methods* (2nd ed.). Beverly Hills, California: Sage

Creating Educational Webmasters: Cyberhood Cooperative Learning Project

Dr. Joseph Bowman, Jr., Assistant Professor,
Department of Educational Theory and Practice
University at Albany, SUNY
Center for Urban Youth and Technology
ED114B, School of Education
1400 Washington Avenue
Albany, New York 12222
518-442-4987-office, 518-442-5008-fax
jb219@columbia.edu

The Internet and the World Wide Web are vehicles to empower human expression, educational development, social change. The concern is how to train and prepare educators and students to use this powerful hypermedia resource. The Center for Urban Youth and Technology (CUYT), in the Department of Educational Theory and Practice, School of Education at the University at Albany, SUNY, New York has been researching the relationship between technology and learning environments in education for several years. The Cyberhood Cooperative Learning Project (CCLP) focused on developing teacher Internet skills for classroom instruction and provided the foundation for teachers to become web designers.

Theoretical Basis for this Research

This researcher has spent several years designing technology based programs in urban settings and researching how the use of technology can improve education and quality of life concerns in these communities. Several researchers and national reports on education and economics in urban centers have expressed these concerns in greater detail. [Kretzmann and McKnight 1993] contend that,

No one can doubt that most American cities these days are deeply troubled places. At the root of the problems are the massive economic shifts that have marked the last two decades. In effect, these shifts in the economy, and particularly the disappearance of decent employment possibilities from low-income neighborhoods, have removed the bottom rung from the fabled American "ladder of opportunity." For many people in older city neighborhoods, new approaches to rebuilding their lives and communities, new openings toward opportunity, are a vital necessity."

It is against this backdrop that I looked at constructionism as one philosophy and mindset that can stimulate change and reform using technology in urban communities. [Papert 1991] provides further clarification when he provides a distinction between constructivism and constructionism. He states:

"We understand "constructionism" as including, but going beyond, what Piaget would call constructivism." The word with the v expresses the theory that knowledge is build by the learner, not supplied by the teacher. The work with the n expresses the further idea that this happens especially felicitously when the learner is engaged in the construction of something external or at least shareable... a sand castle, a machine, a computer program, a book."

I have tried to integrate these concepts because they complement each other and provide a strong research foundation for the work that I am doing with technology and urban centers. I have also extended the constructionist view by including the notion of social construuctionism. [Shaw 1994] contends,

"To social constructionism, the social setting itself is a evolving construction. When the members of a social setting develop external and shared social constructs, they engage the setting in a cycle of development that is critical to determining the setting's ultimate form."

The social setting here refers to urban communities in major cities that can benefit from computers and high speed technology to improve the quality of life and promote educational and economic development.

Using the Web as a Learning Tool

CCLP is a collaboration between the CUYT and the Aframian World Enterprises, a local web designer and several students in our department that were interested in learning about the WWW. Aframian World Enterprises had recently launched the Aframian WebNet that is a repository of African and Afro-American resources (Virtually Afro-centric) on the Internet. The Aframian WebNet also features another large repository (Cosmic repository) of general information about other resources on the Internet. (http://www.he.net/~awe) Mr. Reginald Knox, founder of AWE, stated,

"This site was created because I was concerned about the limited presence that appeared on the Internet as it related to Afro-American and African issues. Once I explored the Internet, I was surprised to find that there is a strong presence on the net, but it was all over the place. I wanted to bring these resources together into a single site. That is how Aframian WebNet was formed".

Designing the Instructional Model

The Aframian WebNet was a large listing of Universal Resources Locators (URL) dropped into the site with no categories and not alphabetized. Some of the URL's did not work, because they had moved to another site or their names were changed. The site was divided into two smaller sites Virtually Afro-centric and Cosmic Resources, which totaled 250 locations. The task was to check all the locations or URL's to determine which ones were operational. Create a list of categories that the sites would relate to. Alphabetize all sites under their selected category, and then write the HTML code to make those corrections. The students were asked to share their experiences of web development, research, and design. Aliya Holmes, a graduate assistant and resident webmaster discussed her thoughts about the project. She stated,

Understanding the African concept that "It takes a village to raise a child", it is our obligation to share this wealth of knowledge with our ëvillage'. Through workshops and demonstrations we attempt to spark an interest for technology in others This venture has been extremely successful in developing and maintaining a site dedicated to the interests of those of the African Diaspora. CCLP attempts to bridge that gap by exposing communities to this new environment. CCLP cares too much to see the children of our villages fall through the cracks."

Discussion

CCLP continues to provide the Aframian WebNet with technical assistance and is now looking for ways to improve the look and feel of the site. CUYT has just completed it's own site, (www.albany.edu/cuyt) and has received a grant from WMHT-Channel 17, a local PBS affiliate to produce a web site about educational video resources in the capital region.

References

[Kretzmann and McKnight, 1993] Kretzmann, J., & McKnight, J. 1993. Building communities from the inside out. Evanston, IL: Center for Urban Affairs and Policy Research, Neighborhood Innovations Network. Northwestern University.

[Papert, 1991] Papert, S. A. 1991. Constructionist Learning. In Idit Harel (Ed.) Cambridge, MA: MIT Media Laboratory.

[Shaw, 1994] Shaw, A. 1994. Neighborhood networking and community building. S. Cisler (Ed), Ties that bind: Building community networks (pp. 134-137). May 4-6, Cuuppertino, CA: Apple Computer.

Acknowledgments

I would like to thank Mr. Reginald Knox for the vision and my graduate assistants, Aliya Holmes, Donovan Smith, Vivian Solis-Fagan, and Juan Vargas for keeping the faith.

Creating Web-Pages in Education on a Limited Budget: Strategies for Success in Cooperative Extension

Robert G. Brannan
Extension Family and Consumer Sciences
Kansas State University
rbrannan@oz.oznet.ksu.edu

Mary McPhail Gray
Extension Family and Consumer Sciences
Kansas State University
mgray@oz.oznet.ksu.edu

The Kansas Cooperative Extension Service provides practical, researched-based information and educational programs to address critical issues facing individuals, families, agricultural producers, business operators and communities in our state. Extension Family and Consumer Sciences mission is to link education with life experience to help Kansas citizens improve their lives, their families, and their communities.

Creating a World Wide Web presence for Kansas Extension Family and Consumer Sciences (FACS) in this era of lower budgets has been a challenge to state specialists committed to maximizing their effectiveness through this new technique. Development of a single homepage through a commercial service would cost between $200-$700 and the resources for this work are not available. However, five strategies for producing low-cost web pages have been identified and used with differing degrees of success to produce FACS Web-pages that enhance the delivery of education and information.

Strategy #1: The "Do It Yourself" Approach

The easiest way for faculty to produce web documents at low cost is to do it themselves. The approach involves two important factors that the faculty member must consider: They must be predisposed to learn HTML and they must have the time to do the work. It is difficult to quantify the amount of time needed to adequately learn HTML and implement web pages; it will vary from individual to individual and from institution to institution.

There are major advantages afforded a faculty member who utilizes this approach. First of all, the choice of what work gets done rests solely with the faculty member and they have total control over content and style. However, there are also several disadvantages. This type of project involves a major time commitment. Also, it is difficult to add work of this complexity to faculty that already feel over-committed. Lastly, maintenance and upkeep of the page will likely be the sole responsibility of the creator.

An example of this approach found within Extension FACS is *The Wonderwise Parent* by Dr. Charles Smith, Extension State Leader for Family Studies and Human Services. Dr. Smith authored his page (***http://www.ksu.edu/wwparent/begin.html***) over a six-week period. He had no previous knowledge of HTML, but was able to develop pages that have received recognition from the Magellan Group, the Mental Health Net, and NBNSOFT.

Strategy 2: Full-time, Dedicated Support Person

Another strategy used to generate web pages is to rely on a full-time, dedicated support person. This technician will be experienced in HTML, and will be ready to create web pages. Faculty will be able to schedule work with the technician on a first-come first-served basis. In order to hire and support such a person, salary, benefits, and equipment have to be budgeted. However, this person has the potential to be spread too thin. Scheduling could become political, which would reduce overall effectiveness.

The College of Human Ecology, which houses Extension FACS, has such a person. However, since he supports the whole college, his workload is enormous. To alleviate this load, he has appointed a staff person within each College unit to support each unit.

Strategy 3: Retraining Support Staff

Redefining the role of existing staff is a continuing process. As staff gain access to new and better equipment, especially easy-to-use HTML editors, this transition can be a major tool for creating Web-based material. The most logical content area for existing staff is work that they already produce. Newsletters, fact-sheets, and educational programs lend themselves to this type of work. However, it may be difficult to process archived material into HTML documents due to time restrictions on staff.

An example of the successful utilization of this strategy is found at ***http://www.oznet.ksu.edu/dep/fnut/newslet.htm***. This newsletter (Foods and Nutrition Digest) is a bi-monthly publication. The staff person responsible for setting this document in Aldus PageMaker can easily convert it into HTML.

Strategy 4: Students as Web Resources

Using students appears to work very well in generating low-cost web pages. Students want to have resume material and producing Web-pages seems to suit this purpose very well. FACS has had great success with utilization of student workers. The use of students allows for project oriented work. The students are highly motivated by the end result of having their name on a web page. The disadvantage of using students is that they have to be trained and may have no familiarization with the subject matter material.

Using minimum-wage students has allowed for the creation of Web materials that were only available in hard-copy. Student volunteers were responsible for the creation of an archive of fact-sheets from Extension Foods and Nutrition dating back to 1993 (***http://www.oznet.ksu.edu/dep/fnut/timely/timely.htm***). Since these fact-sheets are produced on a bi-monthly basis, the departmental secretary will now be responsible for converting these documents into HTML as she creates them. This is an example of how different strategies can be coordinated into an overall strategic plan for designing web-pages.

Strategy 5: Students Work as Part of Curriculum

Students can also be used on a per-credit basis, either individually or within a classroom setting. This strategy can produce material such as class projects, practicum projects, or special problems. Students in this setting have a general familiarity with the material. However, students in a group setting have different levels of technical ability which can present problems.

FACS has had great success with the use of a student practicum involving the creation of web-pages. The outcomes of this practicum were all positive: 1) The student learned HTML and earned three credits for her work; 2) FACS received well-done web-pages at essentially no cost; and, 3) The interaction between the Extension and Resident faculty of the Department of Foods and Nutrition was strengthened.

Summary Comment

Overall, the retraining of existing staff and the use of students appear to be the direction that FACS will lean in the continuing development of Web-pages. It has proved effective and should continue to be effective in the future.

Zippers: A Focus+Context Display of Web Pages

Marc H. Brown
DEC Systems Research Center
130 Lytton Ave.
Palo Alto, CA 94301
mhb@pa.dec.com

William E. Weihl
DEC Systems Research Center
130 Lytton Ave.
Palo Alto, CA 94301
weihl@pa.dec.com

Abstract

This report describes *zippers*, an application of outline-processor technology to the display of Web pages. Zippers allow users to expand and contract selected sections of a document, thereby displaying simultaneously the contents of individual sections of a document as well as its overall structure. Zippers can be implemented either directly in a Web browser or by a proxy (and consequently used by any off-the-shelf Web browser); in either case, no changes to HTML source files are required.

Overview

As people rush to put information onto the World Wide Web, more and more structured documents are appearing. These documents include home pages with lots of hierarchically organized links, papers with many sections and subsections, and other lengthy documents such as books and manuals.

All Web browsers that we know of display each page as a continuous scroll. Unfortunately, it is easy to lose context when viewing a snippet of the whole, and it is hard to jump to arbitrary places in the page. These problems become more severe with longer pages.

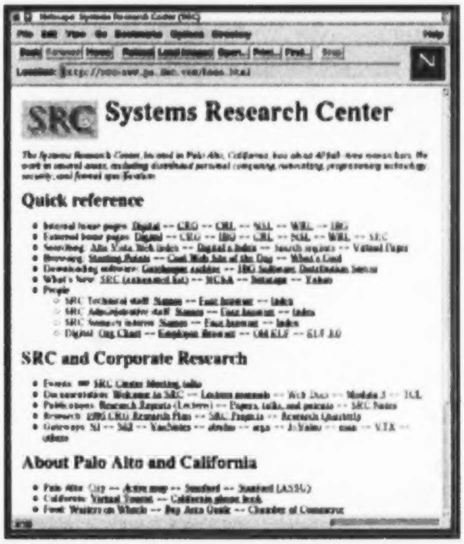
Outline processors provide one well known technique for addressing these problems. Outline processors allow a user to expand and contract selected sections and subsections of a document, thus retaining the high-level structure of the document while also displaying individual sections. In addition, outline processors allow the user to rapidly jump within the document. This report describes the application of outline processor technology to the display of Web pages.

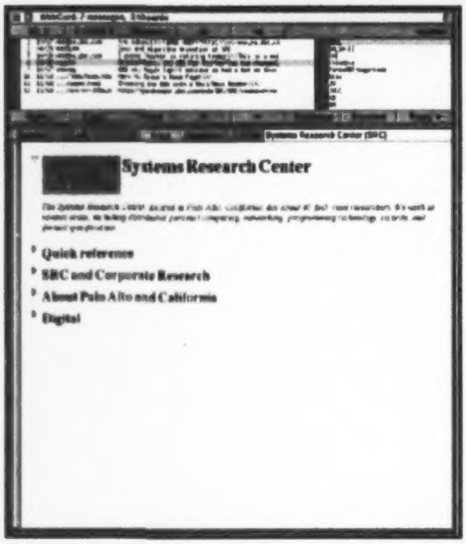
Zippers

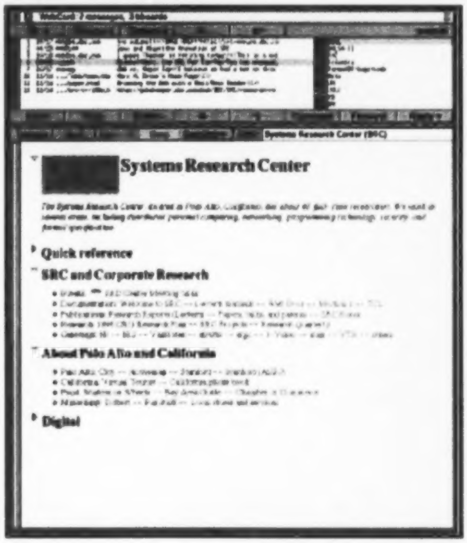
Our display of Web pages infers an outline structure from the heading tags (i.e., H1, H2, ..., H6) in the HTML source. We mark each heading with an icon called a *zipper*. Clicking on the zipper causes the section introduced by the heading to expand and contract. Re-expanding a section causes the state of all subsections to reappear as they were before the section was contracted. However, shift-clicking on a zipper icon causes the section *and all of its subsections* to expand or to contract.

The two sets of screen dumps below shows a Web page using Netscape (top) and using zippers (bottom). The zipper display is within the WebCard browser, an integrated system for handling e-mail messages, bboard articles, and Web pages [3].

The screen dumps below show the SRC home page. The page contains a level-1 heading (a graphic with the letters SRC followed by "Systems Research Center"), some text, and then four level-2 headings. The bottom-left screen dump is the default display with zippers: The single level-1 heading is automatically expanded, thereby causing the text it contains and any subheaders to be displayed. By default, subheaders are displayed contracted. In the bottom-right screen dump, we've expanded the second and third subheaders.

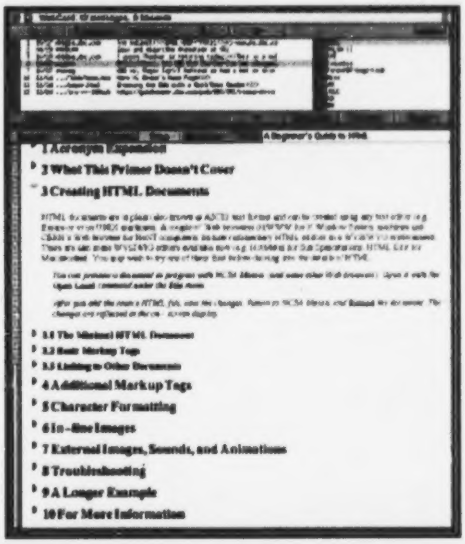

SRC
Systems Research Center
Quick reference
SRC and Corporate Research
About Palo Alto and California


Systems Research Center
Quick reference
SRC and Corporate Research
About Palo Alto and California
Digital


Systems Research Center
Quick reference
SRC and Corporate Research
About Palo Alto and California
Digital

The screen dumps below show a large document, "A Beginner's Guide to HTML" by NCSA [1]. (We edited the page to add section numbers, in order to facilitate the description of zippers.) The Netscape display is 21 screens of information, scrolled to the spot just before section "3.1 The Minimal HTML Document." The zipper displays are scrolled to about the same location. In the bottom-left screen dump, we've expanded section "3 Creating HTML Documents" of the default zipper display. The contents of the section are displayed ("HTML documents \dots\ screen display."), and the three subsections are visible in their contracted form. In the bottom-right screen dump, we've expanded three more sections: "3.2 Basic Markup Tags," and then "3.2.2 Headings" (which only became visible after "3.2 Basic Markup Tags" was expanded), and finally "3.3 Linking to Other Documents."

BEST COPY AVAILABLE



A Beginner's Guide to HTML

1 Acronym Expansion

2 What This Primer Doesn't Cover

3 Creating HTML Documents

3.1 The Minimal HTML Document

3.2 Basic Markup Tags

3.3 Linking to Other Documents

4 Additional Markup Tags

5 Character Formatting

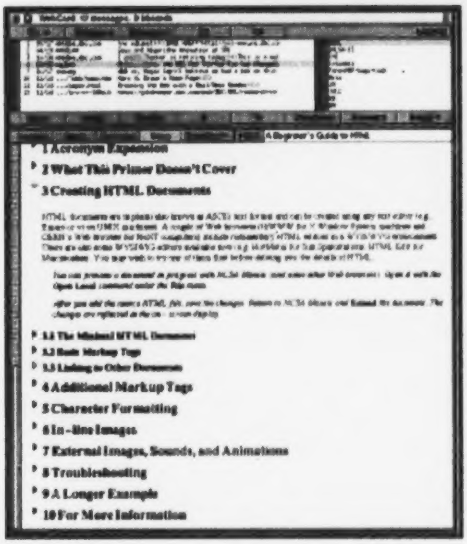
6 In-line Images

7 External Images, Sounds, and Animations

8 Troubleshooting

9 A Longer Example

10 For More Information



A Beginner's Guide to HTML

1 Acronym Expansion

2 What This Primer Doesn't Cover

3 Creating HTML Documents

3.1 The Minimal HTML Document

3.2 Basic Markup Tags

3.3 Linking to Other Documents

4 Additional Markup Tags

5 Character Formatting

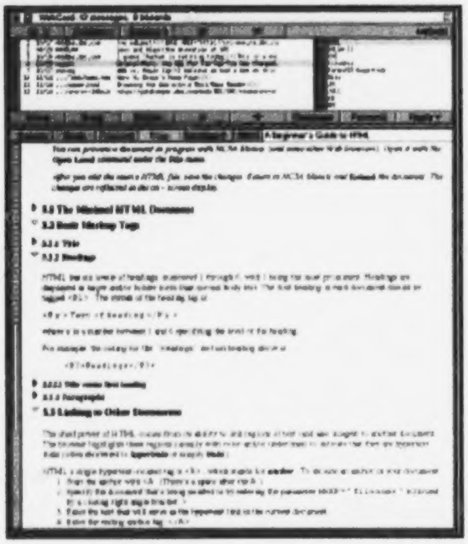
6 In-line Images

7 External Images, Sounds, and Animations

8 Troubleshooting

9 A Longer Example

10 For More Information



A Beginner's Guide to HTML

3.1 The Minimal HTML Document

3.2 Basic Markup Tags

3.2.1 Title

3.2.2 Headings

3.3 Linking to Other Documents

Discussion

Zippers work well on Web pages that use heading tags to indicate the logical structure of the page. Unfortunately, many authors use heading tags to produce particular formatting effects, such as font and pointsize changes, rather than indicate logical structure. On such pages, users simply turn off the zippers and the page is displayed conventionally.

We know of one Web site that has hand-coded zippers [7]. The drawback of hand coding zippers is that it requires server interaction to expand or contract, and an exponential number of files (or a complex script) to generate all states.

Authors of long pages frequently add a table of contents to the top of the page, with entries in the table linking to the various sections. Sometimes, each section appears on its own page, with "next," "previous," "up," and "top" links on each page. Unfortunately, it is easy for the viewer to lose the big picture because the table of contents is not visible while looking at the contents of a particular section.

An alternative to zippers for retaining the big picture while viewing the details is a multi-panel display, such as in SuperBook [2]. One panel displays the table of contents and the other panel displays sections in the document. Clicking on an entry in the table of contents causes the other panel to display the corresponding part of the document. The user can also open and close levels of the table of contents, but the body of a section is always displayed in the second panel. A multi-panel display such as Superbook could be implemented using Netscape's frames, by adding appropriate HTML markups and Java code to a source document.

Another related system is SoftQuad's Panorama Pro [6], which displays SGML documents using a two-panel viewer similar to that of SuperBook. Panorama Pro is also loosely integrated with the Web: The application can be configured as an external viewer of documents whose MIME type is SGML. Clicking on a URL in a document displayed by Panorama Pro will cause Netscape (or some other user-specified browser) to fetch *and display* the URL. Electronic Book Technologies' DynaText [4] is another impressive SGML viewer; it is integrated into Netscape as a plug-in. It's important to realize, however, that Panorama Pro and DynaText are not Web browsers.

Our decision to expand an outline view in place was inspired in part by Pad++ [5], a zooming graphical interface, and in part by the host of successful "context+focus" visualizations developed at Xerox PARC.

We are currently working on a proxy-based implementation of zippers that allows zipper-enhanced pages to be displayed in any Web browser. The idea is simple: A "zipper proxy" runs on the same machine as the user's browser and sits between the browser and the browser's normal proxy. The zipper proxy uses the normal proxy for fetching pages, and modifies the returned HTML to elide material and to include zipper icons. Each zipper icon is linked to a specially constructed URL that is intercepted by the zipper proxy and which allows the zipper proxy to return an appropriate view of the page.

Conclusion

Viewing a Web page using zippers has the nice property of allowing one to see details of parts of the page while maintaining the global context of the entire page. The larger the Web page -- or the smaller the vertical space available for the display -- the more benefit there is to this technique. Putting this technology into a browser or into a proxy seems preferable to hand-coding it into each document.

References

[1] A Beginner's Guide to HTML
`http://www.ncsa.uiuc.edu/demoweb/html-primer.html`

[2] About Bellcore's SuperBook(R) Document Browser
`http://superbook.bellcore.com:80/SB/`

[3] Marc H. Brown.
Browsing the Web with a Mail/News Reader.
Proc. of the 8th ACM Symposium on User Interface Software and Technology, pages 197-198, November 1995.

[4] EBT--Product Information
`http://www.ebt.com/docs/prodinfo.html`

[5] Pad++: Zoomable Graphical Interfaces
`http://www.cs.unm.edu/pad++/`

[6] SoftQuad: Panorama PRO
`http://www.sq.com/products/panorama/panor-fe.htm`

[7] UCI Bookstore Ordering Information
`http://bookweb.cwis.uci.edu:8042/Orders/orders-pt.html`

The VENetIan Virtual Archive Project and a New Approach to Cultural Heritage

Agata Brusegan
Marsilio Editori S.p.A.
Venezia (Italy)
e-mail: marsilio.veniva@tol.it

Alvise De Michelis
Marsilio Editori S.p.A.
Venezia (Italy)
e-mail: demichelis@tol.it

Luigi Giaculli
Marsilio Editori S.p.A.
Venezia (Italy)

Luisella Romeo
Marsilio Editori S.p.A.
Venezia (Italy)
e-mail: luisella@unive.it

Introduction

Assuming that the existence of human society is based on the memory of the past events, the main task of the preservation of cultural heritage has been undertaken by public and private institutions for thousands of years. Nowadays, some of the most important cultural institutions, i.e. libraries and archives, are experiencing an ever-increasing dramatic situation. On the one hand they have a smaller visibility in comparison with other institutions such as museums and on the other they assist at the fast deterioration of the materials they preserve [Brusegan et al. 1994], contemporaneously receiving an amount of requests to access the same materials larger than they can afford. We are strongly convinced that thanks to the electronic media [De Michelis 1994] and the virtual access, the preservation of original documents and their consultation will be thus improved.

The VENetIan Virtual Archive (EP 20638) ESPRIT Project was born out of the need to explore a possible answer to both aspects of the question of archival and bibliographical repositories and the safeguard of the publishing production cycle that originates from cultural heritage. The project aims at creating an on-line virtual archive displaying both archival-bibliographic documents and an editorial collection of journal publications, but it also aims at defining a workbench for the archivist and the librarian, as well as a workbench for the scholars.

The VENIVA Project's Subject and Main Objectives

The main subject of the VENIVA project focuses on the history of Venice in the period ranging from the Middle Ages and the city's defeat in 1797 till the Unification of Italy. Its purpose is to explore and analyse the political, economical and cultural relationships between Venice, Austria and Greece.

The main objective of the prototype is to develop and implement remote access to selected primary cartographic documents preserved in the distributed databases located in Venice (the State Archive and the Marciana Library), Cephalonia (the State Archive of Cephalonia, Greece) and Vienna (the Austrian State Archive) that relate to Venetian history. More precisely, documents and maps belonging to the State Archives of Cephalonia, Venice and Vienna, iconographic collections of sketches, drawings and maps of Crete and other Greek territories preserved in the Marciana Library will be virtually linked in the VENIVA database.

From an applicative perspective, the single historical archives and libraries, once they have been assembled and made accessible as a whole, will constitute a single virtual archive available on the World-Wide Web for the several connected users. The prototype will work as an experimental platform for a new management system of innovative services for libraries and archives that wish to increase the value of cultural heritage by means of remote access.

At the same time, the VENIVA virtual archive will also include and create an on-line editorial collection of original works written by historians that have done research studies on selected primary documents. On this specific subject you can find the 1st title [Viggiano A., 1996] of this collection at: http://veniva.tol.it.

In general the project aims at building an information network of heterogeneous contents ranging from primary sources to critical and historical analyses of the sources themselves. Scholars will contribute to the VENIVA collection with journal articles, conference proceedings, graduation or PhD dissertations.

The VENIVA Virtual Archive's Services and Main Work Environments

We imagine the VENIVA pilot application as an on-line virtual archive that presents a database of cartographic materials and a collection of essays written on Venetian history. At the same time, VENIVA can also be defined as an on-line workbench for users connected to Internet who wish to work with the documents preserved in this virtual archive.

The nature of the VENIVA application is therefore structured into two separated levels which are however strongly interconnected. On the one hand, we have an open environment for users whose main purposes are browsing and reading as well as doing authorial work on the materials available on the VENIVA virtual archive. On the other, we have the publisher's private environment.

The services, the VENIVA virtual archive will offer will deal with the access to archival and bibliographical sources and with document delivery. In addition to this, there will be special services, such as virtual reality reconstructions of historical settings, guided tours and the bookstore where one can purchase the VENIVA publications.

The open environment

Users will have access to the VENIVA platform through Internet. In particular, the virtual archive will appear to them as an usual WWW page. The potentialities that it offers, however, will be different from what is nowadays available on the Web. The exploitation of these potentialities depends on the users' needs and choices. Users, in fact, will either behave as common readers or ask for authorial work to be accomplished.

- The reader environment. Readers of the VENIVA materials can select among three main areas to view: the editorial collection, the primary archival documents and the bibliographic materials.

The reader can either browse the series of publications related to Venetian history or pose a search query and retrieve those items that most interest him. Readers can view the descriptive cards of the selected essays, view and read them, follow the links that connect the author's elaboration to the primary sources he refers to.

Readers will view descriptive cards of the cartographic materials and, if available, can have access to their digitised copy. Also, they can pose search queries according to a selected parameter (for instance, one of the voices of the descriptive card, or through a full text retrieval based on keywords of the reader's choice).

It can also happen that these documents have been analysed or simply quoted in one of the essays of the

VENIVA editorial collection. In this case, readers can follow the link from the primary document they are viewing to the critical analyses that relate to it.

- The authorial environment. Users can be more interactive with the materials stored in the VENIVA database. They will not only browse the documents, read them or follow the connections that are automatically created by the system and the links that the authors of the publications have established; but, in the open environment, users can be authors, too.

The researcher will connect to the VENIVA platform in order to examine the materials he can not otherwise access, unless he physically visits the archive or library that preserves them. In particular, he will be able to search for archival and bibliographic documents, extract information from the publications he has read or the primary materials he has viewed, elaborate them, create links establishing hotwords, add new text and images, save his own work and search results on a personal folder on his desktop, and eventually submit his work to an editorial committee for publication.

The private environment

The main tasks of the publisher -i.e. the subject that offers editorial services with added value in the open environment- can be thus considered. On the one hand, the publisher is in direct contact with both the reader and the researcher that enter the VENIVA archive. On the other, the publisher will communicate with the cultural repositories of primary information -i.e. archives and libraries. If they wish so, the publisher can administrate for them the financial exploitation of their materials. Finally, the publisher will control the virtual archive and its automatic updating of data.

Users can also prepare a project for a possible editorial publication and ask the publisher to produce it for them. They know the main characteristics of the editorial product they wish to buy and can provide precise directions for the publisher. In this sense, users will ask for a personalised product, choosing between different levels of editorial refinement or different supports (electronic or traditional). The publisher will fulfil their requests and will submit their work to the editorial committee that will decide upon the feasibility of its publication on-line.

References

[Brusegan et al. 1994] Brusegan, A., & De Michelis, A., & Ellero, G., (1994). Miners, un progetto per la ipertestualizzazione degli archivi pubblici: il caso dell'Archivio delle Zitelle. Association for History & Computing 7th International Congress, Bologna, IT. 109-118.

[De Michelis, 1994] De Michelis, A., (1994). I Giacimenti Culturali come risorsa fondamentale dell'Europa del 21° Secolo. Archivi & Computer, 1, 84-90.

[Viggiano 1996] Viggiano, A., (1996).Virtual Tour in the Venetian Historical Sources Route: Venice and the Levant (cent. 15th - 18th). http://veniva.tol.it, VENIVA Consortium.

Cyberschool - Singapore Polytechnic Model

Eric Cheung,
Ng Song Cheng,
Ho Lai Har
Dept. of Electrical Engineering
Singapore Polytechnic
500 Dover Road, Singapore 139651
email : echeung@sp.ac.sg

Introduction:

This paper describes a project conducted in Singapore Polytechnic. The project aims to develop tools for "Cyber-school", which uses World Wide Web (WWW) to create, visualise and maintain a uniform definition of objects, scene arrangements, and spatio-location which is consistent across all of the Internet. Several technologies have been invented to handle the scale problems associated with widely-shared spaces, including a distributed server methodology for resolving spatial requests. Virtual Reality Markup Language (VRML) [Name 94] is introduced as a beginning proposal for WWW visualisation.

It seems reasonable to propose that WWW should be extended, bringing its conceptual model from two dimensions, out, at a right angle, into three. To do this, two things are required; extensions to HTML to describe both geometry and space; and a unified representation of space across Internet. With the VRML 2.0 [VRML 96] supports complex 3-D animation, simulations, and behaviors by allowing Java and JavaScript programmers to write scripts that act upon VRML objects. It is shown that the usage of VRML is a very efficient and powerful method for the transport of information over the Internet.

Objectives

(a) To provide tools for different organisation (industries) to implement VR to the Internet by means of VRML

(b) To provides a structural approach in describing a virtual reality scene.

(c) To develop a prototype model call "Cyber-school", to provide information about Polytechnic.

(d) To provide a true desktop VR tools benefits to Education

Implementation

Basically there are 3 ways to create a VRML world:

i Create and edit a VRML text file by hand

ii Use a conversion program to convert an existing non-VRML 3-D file to VRML

iii Use an authoring package to create models and position them within a world

This project will be using the combination of the 3 ways to create the VR world for user to walk & browse through.

Phase 1 of the project is to create a Virtual Walkthrough model of the Campus (Dept. of Electrical

Engineering). In total there are more than 20 laboratories in the department itself. All the labs set up will be transformed to the VR model and later on can connect to the headgear and gloves to control the movement. The user will be able to walk in to different labs and browse through the whole lab. Multimedia will also be used in the model. User will be able to watch some of the experiment and even ask question about the subjects related to the laboratories and get some advice from the lecturer. And they can even launch a video clip about the lectures related to the lab. The proposed package will be housed in the SGI Oynx Machine and connected to a LAN for trial run.

Phase 2 & 3 of the project will be implemented to all the different departments within Singapore Polytechnic. And later on will the programs will be put into the Internet Server in Singapore and allows people from all over the world to view and visit Singapore Polytechnic.

Scene Graphs

Figure 1 and 2 show some of the simple model created for the project. The user will be able to walk through the lab and by clicking certain position, different types of information will be shown to the user.

Benefits to the Students

The set-up used for the proposed project will provide a good opportunity for the staffs and students to learn the concepts of Internet & Virtual Reality Language. Data can also be translated to other formats (color, intensity etc..) as it generated, and then easily reviewed or for further analysed. This can be particularity helpful for students with physical or perceptual disabilities. Students can have access to apparatus which in real life is too expensive, too delicate, too big and too dangerous to keep in classroom.

Benefits to Industry

With the growing awareness of the Internet, many companies like to explore the possibility offering services or market their products on the Internet. The proposed project can be used as a demo to the industry on putting VR in the Internet. With the prototype capability, different industry can see the benefits offering VR in the Internet. They also be able to evaluate the possibility of putting their products or services in the Internet. The system can be easily adapted to Industry by means of VRML. At present, seventeen companies and organisation had already shown their support for VRML-based 3D graphics on the World Wide Web.

Summary

The rapid, nearly exploding growth of the Internet & WWW has already been shown to the general public how powerful this new communication concept is. In Singapore Polytechnic, we also believe the similar growth for VRML. With more and more vendors supporting VRML & 3D in the Internet, it could become a standard format for exchange information and engineering applications.

Acknowledgments

We like to thank Dr. Tan Tin Wee, National University of Singapore for providing VRML

browsers and advice on creating VRML object. The work is partly supported by the TOTE board funding by Economic & Development Board, Singapore

References:

[Lee 94] T Berners-Lee VRML- the building space

[VRML 96] http://vrml.wired.com/vrml.tech/vrml0-3.html

US Information Superhighway: Architecture and Applications

H. Michael Chung
Department of Information Systems
California State University, Long Beach
Long Beach, CA 90840-8506
USA
Internet: hmchung@csulb.edu

1. Overview

It is a misconception that the information superhighway will develop like the federal highway project. While the roads were primarily funded by the federal government, the government is not going to inject the financial resources for the information superhighway. Instead, the government will change the ground rules for the industries to lead the National Information Infrastructure (NII).

A bill passed early this year will bring broad changes in the telecommunications area. For example, long distance phone service will also be provided by regional Bell phone companies, which could lead to lower long distance rates. At the same time, local phone markets will be open to traditional long distance as well as cable TV firms. Since they need to pay local phone companies in order to connect into the existing network, local rates may rise. Cable rates will go up due to the lifting of rate regulation. National TV station ownership cap will be raised. Also, cross ownership between cable and telephone companies will be allowed so that new alliances will emerge.

In this setting, the primary driving force behind the NII is not the new technology nor new regulation. It is the demand for new products and services. Namely, the government is not really leading the NII effort except in several public areas. Rather, the market is shaping the NII amid the short term inefficiencies and the deficiencies in nationally visible showpiece monumental outcomes.

Furthermore, by the year 2000, over twenty percent of all Internet access devices will be non-microcomputer machines. This is another misconception that the NII is always related to PC's and its compatibles. Consumer demands and citizen's expectations are the core guideline of the direction of the information superhighway. Expectations can be met and also demand can be created. To satisfy this assertion, capital should be consumed. However, the industry will not move without being convinced that the information superhighway will guarantee being the profit center for the participants. In other words, business applications are the most critical area that will motivate the industry.

2. Prospects

After witnessing the rise and fall of somewhat intended hypes, consumers and the industry clearly see the real challenges and the opportunities that lie ahead on the path of constructing NII: an understanding that the information superhighway has been less a blueprint and more an inkblot test of multiple architects from different firms.

Highways are a particular thing that has been designed to fulfill a clear need to link distinct points. On the other hand, NII will not be built on the blue print. It will evolve through the practical merging of the hardware, software, circuits, and applications. The results will be an invisible, seamless, and dynamic web of networks, and information resources. Naturally, this may lead to the technological incompatibilities in engineering, opaqueness in application transparencies for consumers, and short term economic inefficiencies in resource allocation and utilization. For instance, cruising the information superhighway versus crawling the road exemplifies an engineering goal the internet has to achieve. In addition, an acid test that questions whether the internet changes the nature of everything, or it is a new way to do the same old stuff symbolize the way the internet is to aim at.

For the engineering side, an example may be an optical network layer with cross-connect and add/drop functionalities. Eventual optical multiplexing may be able to promote high bandwidth utilization. For the management side, distributed systems that comprise of diverse types of vendor hardware and software have to be transparent. Management effectiveness accompanies an issue of standards in need of interoperability. Some of the key technology service areas are private line, frame relay, Integrated Services Digital Network (ISDN), Asynchronous Transfer Mode (ATM), and Satellite.

In constructing NII, particular attention needs to be paid on the global infrastructure, standards, and products and services. It is necessary to consider some of the critical factors in designing the NII: protocols or applications that require extensive back and forth transmission of small packets. Network segmentation to enhance the performance in interfacing the frame relay with X.25. Consideration of remote procedure calls over traditional queries wit respect to the graphical user interface. Effects of distributed versus centralized control of business on the amount of data flow over the circuit.

3. Pilot Cases

In late 1980's, the City of Cerritos in California tried to deliver a large library of movies to residents within thirty minutes. However, many residents balked at the extra cost of services. Menu of concocted products like video on demand, especially on a fee basis may still not be welcome for both consumers as well as for those who have to bear with the high cost of installing fiber optics. Last year, PacTel decided to slow down an year long effort of upgrading the city of Los Angeles communication network to provide the video dial tone quality, and instead will resort to wireless cable technology.

In late 1990's another California project at Playa Vista is underway to address the city wide integrated information infrastructure to accommodate both analog and digital communication media along with the high tech entertainment industry.

A noteworthy example of success nearby the Playa Vista project location is a net drama, The Spot. The author participated in producing a telecommunication technology education project that will be disseminated via the Internet. Many producers of these types of products are gathered around West Los Angeles seashore and earned the name of being Silicon Beach.

In addition to the entertainment and education industries, Web based sales of a retail product as exemplified by Hot Hot Hot encourages new types of businesses. Moreover, personal finance and banking, and yellow page markets are apparently some of the good examples worth watching. Integrating fixed and mobile telephony services may be the one that needs more attention. In fact, it is one thing to have a good fast hardware or highway and it is another to utilize such capabilities and add values appealing to consumers.

4. Lessons

Considering the characteristics of information consumers expect, content is the first consideration while the quality and value of the content may be even more essential. Secondly, consumer interface, often called graphical user interface and accessibility are critical considerations. Finally, the cost of such information service should be weighed.

Collaborative Design: A Cognitive Approach To Information Resources

Alex Cuthbert
Education In Math, Science, & Technology (EMST)
University of California, Berkeley; Berkeley, CA 94720-1670 USA
Email: alx@violet.berkeley.edu http://www.kie.berkeley.edu

These studies address two questions: (a) Does an activity structure that encourages iterative refinement help expand the problem definition of open-ended design tasks involving Internet resources? and (b) What types of online resources help students locate information on the Internet? By addressing these questions using the Scaffolded Knowledge Integration (SKI) framework [Linn 1995], this research provides suggestions for improving the design of tools that facilitate knowledge construction and integration in online environments. These studies extend research in the areas of collaborative design, pragmatics and communication, category and concept representation, environmental science education, and information retrieval. The activities took place within the Knowledge Integration Environment (KIE), a computer-based learning environment combining on-line tools such as a student notebook, a multimedia discussion forum (SpeakEasy), and on-demand cognitive and procedural guidance.

In the first study (N=131), eighth graders searched the World Wide Web for information relevant to a design task involving the storage of heat and the regulation of temperature. The activity consisted of surveying selected Internet sites related to energy conservation and environmentally-sensitive housing design, searching the Internet for evidence to support original designs, and formulating a final report integrating drawings of a desert house with principled evidence for design decisions. The majority of the students were successful at searching, locating, and incorporating Internet evidence into the design process. The degree of principled reasoning behind the design decisions varied considerably and did not correlate with the success of the search activity.

In a second experiment (N=132), a collaborative search page was added where sites selected by students were automatically added to a publicly-accessible and searchable Web form. Log files were generated to track student activity patterns, queries, and comments about Internet sites. The new activity design was successful in expanding the students' definition of the problem as measured by the range of features considered. The publicly-accessible and searchable Web form containing student-located sites helped scaffold those students that had difficulty locating Internet material on their own. Extending earlier work in the area of information resources and design tasks [Cuthbert 1996], this research suggests that the mechanisms that help students critically analyze marginally relevant data and develop cogent, principled arguments are tied to the ability to conceptualize the problem as opposed to the ability to generate queries to narrow the search space.

Methodology

The eighth graders involved in this study had taken a semester of science using the Computer As A Learning Partner (CLP) curriculum. Students were allowed to select their partners for the activity. The activity called "Desert Houses" was an end-of-term project designed to allow independent work linking principles about insulation and conduction that were discussed throughout the semester. The participants worked in pairs on Macintosh LC II computers using Netscape, Microsoft Works, and the KIE environment . During the activity, the participants examined actual Internet sites, prepared sites, and excerpts from labs completed earlier in the semester (e.g., a "heat bars" lab that measured the conductivity of various materials). They searched for relevant Internet sites by generating keywords as a class and then entering words into a form that fed into a search engine (i.e., http://www.excite.com). In the first experiment, students had access to an on-line thesaurus. In the second experiment, they had access to sites located by other students through a searchable Web form. In the second experiment, the activity checklist included: initial design, survey evidence, synthesize evidence, online discussion, search episode, heat flow analysis, redesign, and final report. Clinical interviews with six students provided an in-depth look at the conceptual understanding behind the designs.**Results & Discussion**

The majority of the students were successful at searching the Internet for relevant sites, saved evidence to the Netbook, and incorporated that evidence into their final reports. The pattern of search activity that emerged was more characteristic of perceptual search patterns directed by the stimulus of the display rather than the logical deployment of a systematic search strategy. Access to the sites located by other students provided relevant information for students that had difficulty locating material on their own . However, access to information did not necessarily lead to critical thinking about the sites. The addition of a Revised Design Worksheet and Heat Flow Analysis Worksheet had little impact on the selection of features considered for the dwelling (e.g., windows, size of rooms, orientation of house, type of insulation, etc.) but did reveal internal conflicts within the students' explanatory framework for the temperature differences at different times of day. This cognitive dissonance occasionally resulted in a reevaluation of the design or reassessment of the explanatory mechanisms for describing heat flow. Initial designs reflected a strong bias towards the examples used to illustrate heat flow and towards the selected sites that were visited early on in the project. Students generally retained the framework of their initial designs and used evidence as a "proof of existence" for features of a house rather than as an argument for selecting one design over another. Principled reasoning about the dwelling design came late in the process in either the Heat Flow Analysis worksheet or the Final Report. The use of prototypes as a starting point for the design rather than principled reasoning about temperature regulation increases the importance of the examples and the accompanying metacognitive prompts built into the software environment. This reliance on selected examplarary sites appears in both the house design and the final report where Internet sites are linked into the final write-up.

The tendency to not revise initial designs suggests that the sequence of searching for evidence to support existing designs should be reordered. Searching for evidence needs to be linked more directly with the conceptualization of the problem if the located information is to have a significant impact on the project design. A number of approaches to encouraging a more dynamic and expanded problem definition can be drawn from other KIE projects such as critiquing other students' designs or developing simple prototypes highlighting the scientific principles involved. Linking the criticisms of existing designs and components of the prototypes with Internet evidence may help model the process of critical thinking that needs to occur for students to successfully complete similar design projects.

The process of using Internet resources to support design decisions seems to have two phases:

(a) locating evidence and (b) using evidence. Tied to the problem of locating and using evidence is the conceptualization of the problem [Cuthbert 1996]. This conceptualization occurs rapidly although the resulting framework appears to be fluid enough to encompass potentially conflicting elements. The process through which these elements are sorted, critiqued, and supported has been the focus of this research. The change in activity structure from the first to second running of the intervention successfully expanded the problem definition for the majority of the students. This expanded problem definition helped students generate a wider range of keywords. Useful sites were automatically added to the publicly-accessible page of relevant Web sites thereby helping the lower-end students collect potentially useful information. The fact that this information was not used to revise existing designs suggests that information resources have significantly different effects on the problem solving process depending on whether they are introduced during the conceptualization or realization phase of the activity. Since each of these two phases of the activity overlap, intertwine, and can be further decomposed, the effect of informational resources on problem solving in general is particularly complex. The results of these experiments begin to show the effect of information resources during the realization phase of the activity. Further research is needed to uncover the potential role of information resources during the other phases of a design activity.

References

[Cuthbert 1996] Cuthbert, A. (1996) How do students use Internet evidence to complete projects? From exploration to directed search. Poster presented at the Annual Meeting of the American Educational Research Association (AERA), NY, NY. (See http://www.kie.berkeley.edu/people/alex/pubs.html)

[Linn 1995] Linn, M. C. (1995). Designing computer learning environments for engineering and computer science: The Scaffolded Knowledge Integration Framework. Journal of Science Education and Technology.

A Web Based Decision Support System

Alanis David, Ph. D. (dalanis@campus.mty.itesm.mx)
Ramses Herrera (jherrera@campus.ruv.itesm.mx)
Information Systems Department
Instituto Tecnológico y de Estudios Superiores de Monterrey
Sucursal de Correos " J "
Monterrey, N. L. 64849 MEXICO.
Phone: (528) 358 2000 Ext. 4591, 4088

Distance Education is one of the most important interest areas for the Instituto Tecnológico y de Estudios Superiores de Monterrey (I.T.E.S.M). I.T.E.S.M. is a 26 multicampus university located in Mexico.

In the last few years, I.T.E.S.M. has developed an extended distance education program (recently called Virtual University) that includes: Live satellite classes, Multimedia supported classes, WWW supported classes and Videoconferences.

The Virtual University uses some of the most important educational schemes because of its coverage and complexity. In this scheme, the classes are produced and transmitted from two cities: Monterrey and Mexico's State. Occasionally, some foreign universities like Carnegie-Mellon, Waterloo and Thunderbird, participate as coproducers.

Remote classrooms form the rest of the educational scheme. At this time, beside the 26 remote classrooms in I.T.E.S.M. exist more than 21 remote classrooms in Mexico (other universities and private organizations) and nine in foreign universities (USA, Colombia, Venezuela, Costa Rica, Ecuador).

The Virtual University's main goal is to apply *avant-garde* information technologies to raise education quality and to improve the teaching-learning process.

Modern telecommunication technologies have been implemented to avoid geographic barriers; for example, all of the classes use: Live television transmissions, On-line Remote Interactive System (SIR in Spanish) that allows students and teachers to keep real time communication during the classes, Telephone and fax, and Off-line communication (email and netnews).

According to Virtual University's main goal, teachers are redesigning their classes so they can be supported by emerging technologies like WWW.

An example is a Web Based Decision Support System designed for MBA's. On this system, the students have links to pages where they can get support on decisions refered to their homeworks, exams, etc..

The students grant access to the links with a password supplied by the teacher at the beginning of the course. After getting access, the students can get charts of their individual performance on exams, homeworks, papers, etc.. Another important function of the system is that every student can get charts of performance, classified by: campus, country and classroom.

As the class is divided on parcial notes, the students can introduce some test values for exams, papers and homeworks in order to simulate the notes they could get. After testing some values, the students can chart the results (bar, pie, line) and take a decision that helps them to achieve those notes.

One important goal of the system is to give the students as many information as possible so they can improve their decisions.

Kids as Global Scientists: Promoting a Web of Collaboration Through Internet-based Curricula

Holly Devaul
Nancy Butler Songer
School of Education
University of Colorado
Campus Box 249
Boulder, CO 80309-0249
kgs@spot.colorado.edu
http://stripe.colorado.edu/~kgshtml/Home.html

The *Kids as Global Scientists* program capitalizes on the Web's communication potential to bring together educators, software developers, and scientists in a collaborative effort to provide a rich and meaningful educational experience for middle school students. Developed by Dr. Nancy Butler Songer and colleagues at the University of Colorado, Boulder, our work is supported by the National Science Foundation's Networking Infrastructure for Education and a 1995 Presidential Faculty Fellowship. We can be found on the Web at http://stripe.colorado.edu/ ~kgshtml/Home.html. *Kids as Global Scientists* (KGS) is a new kind of Internet-enhanced science curriculum designed to encourage middle school student inquiry and research. Students use visualization and telecommunication technologies to learn about science both locally and through interaction with peers and resources worldwide. Capitalizing on the power of communication via the Internet, students make contact with their peers and professionals and learn to access the ever-growing volume of information available on the Web. Through guided activities, as well as taking advantage of current weather events as they happen, students learn to integrate multi-faceted pieces of information into a richer understanding of atmospheric science. Since 1992 approximately 4500 students in over 60 world-distributed locations have participated, from such far-ranging locations as Finland, Australia, Hong Kong, Brazil, Scotland and across the U.S.

The KGS research and learning philosophy exemplifies collaborative efforts in many and diverse ways. First, our pedagogical approach embraces the idea of group work; both within the classroom and beyond it's borders via the Internet. Students access real-time imagery, exchange data, and discuss the implications of these multiple representations of data with each other and with a volunteer on-line scientist matched to their site. This kind of beyond-the-classroom collaborative learning (student-to-student, student-to-scientist and students interacting with real-time data and imagery) has previously been absent, or difficult at best to integrate into a classroom curriculum. Integrating technology and Web access into the curriculum is the vehicle to take this step. Our research indicates that the connections that are made via this technology enhance motivation and participation in many students, including females and other groups of students often alienated by the traditional classroom approach. In addition, these connections can serve as a means to encourage the development of rich scientific explanations, a concept we call "interactive knowledge".

Beyond the execution of KGS in the classroom, we provide a framework for participating teachers to work collaboratively via an electronic listserv. The listserv was initiated in 1996 and has thus far, been successful in providing a forum for teachers to share ideas, frustrations and experiences, as well as provide support for one another. The teachers participating in KGS have a wide range of experience with technology and with implementing alternative curricula such as ours. Providing a mechanism to promote support and assistance has proved helpful to the less experienced teachers, and empowering for those able to share their knowledge.

We have enjoyed and cultivated a productive collaborative relationship with our local scientific community, who have been extraordinarily helpful in reviewing our curriculum for content accuracy, in addition to participating in teacher workshops and other professional events. The Boulder Valley School

District has been supportive of our efforts, allowing teachers the flexibility to modify their larger curriculum to mesh with KGS, and in supporting technology in the classroom via a sister project, Boulder Valley Internet Project. In this regard, we have found that the nature of our research has brought us into contact with a wide range of groups; effective collaborations with these diverse groups has allowed us to develop and implement our project with ease.

KGS, now in it's fourth year, is extending its collaborative scope to formally include software developers from the University of Michigan's Weather Underground (BlueSkies team; http://groundhog.sprl.umich.edu/), and scientists and educators from The Concord Consortium (http://www.concord.org/) Project GREEN (http://www.igc.apc.org/green), and the Global Hydrology and Climate Center in Huntsville, Alabama (http://wwwghcc.msfc.nasa.gov/). We believe that upfront collaboration with software developers to produce age-appropriate, effective interfaces will enhance the learning potential of our curriculum and similar technology-centered curricula. In addition, by combining development efforts with other educational researchers as well as content experts, we can mesh the best pieces of our individual strengths to produce a better product, both educationally and technically. A pilot run of our new curriculum, entitled One *Sky, Many Voices: Hurricanes ë96*, is scheduled for October of 1996.

We look ahead towards exploring new areas, including integrating real-time multimedia conferencing software with resource-sharing features. This would allow us to bring together regional schools and scientists to dialog about a current weather event as it happens; sharing images and text in real-time. Other developments include enhancing BlueSkies for student data collection, integrating data analysis, graphical interfaces and messaging into a single package, allowing students to share not only text, but graphical displays of their own design.

In summary, we have found that the *Kids as Global Scientist's* Internet-based curriculum has fostered a wide variety of collaborative relationships, including student learning environments, teacher professional development, and has worked to bridge the gap between scientists and educators. Our current research approach has expanded to allow for collaborative efforts among software developers, educators and content specialists in order to produce a more comprehensive and effective learning tools for middle school students.

HERODOTUS: An Educational Site as an Integral Part of a Study Book

Sjoerd de Vries
Faculty of Educational Science and Technology
University of Twente
The Netherlands
vries@edte.utwente.nl

Henk Donker
Faculty of Educational Science and Technology
University of Twente
The Netherlands
mp_donker@edte.utwente.nl

Introduction

In the Netherlands there is an on-going process of restructuring education. Secondary education is now being restructured into a basic level (the first two/three years) and an upper level (the last two/three years). The leading idea behind the upper level is the so-called 'study home' concept. The intention is that the learners become 'successful independent constructors' of their own study processes instead of 'successful dependent victims' of presented instruction. The first secondary schools start in 1996-1997 with their 'study home' and educational publishers are working on new appropriate study books in order to make the 'study-home' concept a success. To reach for success, publishers consider the use of educational software as an integral part of these new study books. Such educational software is expected to make use of the rapid evolving technologies like hypermedia and the Internet.

Educational Software as Integral Parts of Study Books

In co-operation with an educational publisher we, this is the division Educational Instrumentation Technology of the Faculty of Educational Science and Technology, University Twente, the Netherlands, searched for an acceptable design of the desired educational software. To be acceptable, this design has to function for the next years as a blueprint for implementation and maintenance of educational software products, which products should serve as an efficient and effective integral part of their new study books. We introduced the concepts Interactive Study System and Interactive Study System Environment to refer to these educational software products.

Basically we see an Interactive Study System (ISS) as a software system consisting of study tools, educational resources, and if needed additional materials. This system is an integral part of a study book. An ISS can be extended by generic software tools, like communication and information editing tools.

- Study tools can vary from drill and practice tools, tutorial tools, simulation tools, concept mapping tools, study workplaces, e.g. Study tools are software tools that are particularly designed to enhance individual or group-based study processes.
- Educational resources refer to the multimedia information content matter. The central educational resource is named the CourseBase. A CourseBase can be seen as an object-oriented database containing Instructional Multimedia Information Content Constructs (IMICC's). An IMICC is a construct consisting one or more Multimedia Information Content Primitives (MICP's) designed for instructional purposes. MICP's are basic multimedia information content units serving one goal-directed action, such like identification, relation, presentation or question.
- Additional materials refer to information and/or tools that support study processes but which are

not offered in an electronic way.

- Generic software tools are tools that are not particularly designed to enhance study processes. Examples of such tools are communication and information editing tools. Communication tools can be used to enable communication between students, between student(s) and instructional managers, information providers, e.g. Information editing tools can be used to edit documents, spreadsheets, tables, e.g.
- The main goal of an ISS is to enable an interactive study process between a user of the system and other users and/or between the user(s) and educational resources.

However to use and maintain an ISS appropriate tools are needed. To refer to these tools we use the concept Interactive Study System Environment (ISE). In order to speak of an environment, three conditions have to be met:

- *Full support* of task performance by means of instruments and Electronic Performance Support in each phase of the lifecycle of a product, in this case the lifecycle of an ISS;
- *Flexibilisation of task performance,* the user decides about where, when, and how tasks are performed;
- *Integration,* the user has a consistent and uniform view on the applied systems.

An ISE is seen as a software system for (re-) design, use and evaluation of ISS's.

In co-operation with the educational publisher, we designed and implemented a prototype ISS named Herodotus. Herodotus is an integral part of the new history study book published by the involved educational publisher, which book is intended for the upper level of secondary education in the Netherlands. We consider the prototype Herodotus a first step to the realisation of an ISE.

The Design of Herodotus as an Integral Part of an History Study Book

A main interest in our design of Herodotus is the application of the 'WEB technology'. There are two reasons for that. First, because of the openness of the WEB technology. It offers opportunities to enhance the flexible use and maintenance of ISS's. Second, the WEB offers three basic functions, which are expected to be essential for learners in order to practice becoming 'successful independent' constructors of their own study process. These three functions are: communication, co-operation, and information gathering. These three functions are essential in a wide variety of study processes.

There are two ways of looking at the use of the WEB technology in relation to the design of ISS's. The first way is to design ISS's based on the WEB technology. This implies that an ISS can be seen as an educational site on the WEB. The second way is to consider an educational site a sub-system of an ISS. This implies that users of an ISS, also have the opportunity to make use of an educational site, which site is specifically designed as a sub-system of an ISS. We have chosen for the second way in order to design Herodotus. The main reason is that the need for interaction is a basic principle for ISS. The way the WEB technology allows for interaction (For instance 'Java') is rapidly evolving, however we considered such opportunities for interaction not yet 'stable' enough.

In our presentation we will describe the design of an ISS and in relation to this Herodotus. We describe the main design principles and the architecture. Special attention will be given to the way we integrated Herodotus into the history study book. In addition, we describe the user interface of the educational site which is a sub-system of Herodotus. Finally, we give remarks about further research in the design of ISS's and more specific about educational sites as integral parts of study books or curricula.

Knowledge Explorer Centre: The Web as a Multimedia Publishing Medium

Greg Dixon
The Dixon DesignWorks, Canada
gdixon@dsoe.com

Introducing the Knowledge Explorer Centre

In response to the phenomenal interest in the World Wide Web, the Dixon DesignWorks has changed its view of the Internet from that of promotional medium to that of publishing medium for multimedia educational content. The result is the KNOWLEDGE EXPLORER CENTRE, recently launched on the World Wide Web (www.dsoe.com/explore). This paper examines the benefits, challenges, and viability of using the World Wide Web as a multimedia publishing medium.

A Brief History

The Dixon DesignWorks developed an English language tutorial program called *English Structure and Style* in the 1980's. *English Structure and Style* helps secondary, college, and adult students learn about grammar and style elements. A few hundred schools, school boards, and colleges have licensed the program for use in their language labs.

The Dixon DesignWorks converted the content of *English Structure and Style* to Windows hypermedia format and published it as a shareware program called *Exploring English*. *Exploring English* was the first in the planned Knowledge Explorer Series. Other titles planned for the series included *Exploring English as a Second Language*, *Exploring Literature*, and *Exploring Libraries*. Market research indicated high interest in *Exploring Literature*, followed by *Exploring Libraries*. The results of the market survey, plus exciting new developments on the Internet, have led to the creation of the KNOWLEDGE EXPLORER CENTRE on the World Wide Web.

Objectives of the Knowledge Explorer Centre

The KNOWLEDGE EXPLORER CENTRE is an innovation in Internet publishing. Some of the main objectives are:

- To share educational materials to help students achieve success.
- To encourage others to explore specific areas of knowledge.
- To provide a forum for people of varied backgrounds to share their knowledge and enthusiasm for a wide range of subject areas.

Current Knowledge Explorer Titles

Each main area of the KNOWLEDGE EXPLORER CENTRE provides a multimedia overview to a subject area with pointers to other resources available on the Internet and on other media.

- *Exploring English* provides a webmedia tutorial on the English language for adults and young adults.
- *Exploring Gardens* provides information and photographic images of plants and garden topics.
- *Exploring Libraries* introduces people of all ages to the resources found in modern libraries.
- *Exploring Literature* provides an introduction to the rich world of the written word.

- *Exploring British Columbia* provides a fun introduction to the natural wonders of British Columbia.

Benefits of Publishing on the Internet

The original intention for creating a World Wide Web site was to promote individual titles in the Knowledge Explorer Series, each offered as separate multimedia software packages. However, we shifted our view of the Internet to that of a multimedia publishing medium for the following reasons:

- Natural translation of the hypermedia approach from Windows Help format to webmedia format.
- Instant distribution to millions of Internet users.
- Hardware and operating system independence.
- Increasingly rich multimedia platform.
- Low development and publishing costs.
- Opportunities for world wide collaboration.

Challenges to Publishing on the Internet

There are a number of challenges that may limit the effectiveness of Internet resource centers such as the KNOWLEDGE EXPLORER CENTRE. Most are cultural and not technical.

- Attracting the people who could benefit the most.
- Countering the general perception of the Internet primarily as an entertainment vehicle.
- Regaining the interest of many who have searched for quality educational content without success.
- Finding specific sites on a network which by nature is not organized.
- Wandering through too many lists of lists leading to sites without content.
- Funding the development of quality content.

Sponsorship Funding Model

The KNOWLEDGE EXPLORER CENTRE was launched without any government or corporate funding. Its continued growth will depend on the level of support from the people using the material and from sponsors. Individuals and organizations that find the material useful are encouraged to show their appreciation by contributing to the Center. Companies and organizations looking for public relations and marketing exposure are invited to sponsor the KNOWLEDGE EXPLORER CENTRE in general or to sponsor specific areas such as *Exploring Gardens*.

Advertising space is available in a linked area called the EXPLORER TRADING POST. Advertisers associate their products and services with specific areas of the KNOWLEDGE EXPLORER CENTRE.

Future Plans

Assuming a positive reception from the Internet Community and funding support, the KNOWLEDGE EXPLORER CENTRE will continue to evolve in the following areas:

- Expansion of current areas according to interest and support.
- Expansion into new subject areas such as *Exploring Art* and *Exploring Ecology*.
- Publishing in other media such as books and CD-ROM.

Conclusion

Online multimedia resource centers such as the KNOWLEDGE EXPLORER CENTRE are emerging as important sources of knowledge and inspiration on the World Wide Web. The rich suite of multimedia tools and a staggering potential audience make the Web a natural vehicle for publishing educational and exploratory materials. The success of such sites over the next year will indicate the viability of the World Wide Web as a publishing medium for multimedia educational materials.

Design for Web-Based Learning

Philip Duchastel
Program in Education and Technology
Nova Southeastern University
Fort Lauderdale, USA
http://www.nova.edu/~duchaste

Sue Spahn
Program in Education and Technology
Nova Southeastern University
Fort Lauderdale, USA
http://www.nova.edu/~spahn

"The ways of handling information that work well in 'Old Media' (print and broadcast) do not always translate gracefully into new media environments." - McAdams (1995)

One of the main uses of the WWW (the Web) is informal learning through browsing. The Web is also used for formal education (particularly through various distance teaching programs that offer credit for following particular courses of instruction), but by far its greatest use is informal learning, in which people access and learn from information made available by a whole host of organizations and individuals within a non credit-granting structure. This on-line informal learning may augment more formal off-line educational efforts, but it more typically takes place simply out of the shear enjoyment of learning.

The great majority of information made available on the Web is not specifically prepared for learning purposes. Rather, it is more often developed with a view to interest and inform a general and varied audience. It is in this respect akin to public communication products and less aligned with instructional products, even though it often finds itself at the interface of both realms. From this ambiguity arises the design problem for web-based learning materials. This paper highlights some of the principal design issues involved and proposes the outline of a design methodology that can help developers in their work of preparing web-based learning materials. It is in this respect but an initial effort, but may well encourage further thinking in this important area of Web development.

Two perspectives in evolution

The Web can be seen from two perspectives: as a vehicle for the distribution of resources, and as a medium of expression-representation with its own specificity. The distribution perspective emphasizes the Web as a network infrastructure that has the great advantage of flexibility and low cost. A simple example is the case of university course materials (syllabus, readings, lecture notes, etc.) being placed on the Web and made available to students. The fact that Web information is readily updatable and that users always access the latest version provides tremendous flexibility by eliminating the outdatedness of printed materials; and the electronic nature of the medium provides cost savings compared to traditional paper-based alternatives.

What are being distributed at the moment on the Web are multimedia documents. However, the advent of Java and other similar languages enable true interactive instructional transactions on the Web and will likely totally reshape the field of instructional software. *What this means for the design of such software and how it will affect the larger picture of web-based learning is explored in the paper.*

The expression-representation perspective emphasizes the Web as a novel, distinct form of publication of informative materials that does more than simply replicate other forms. *As we see at greater length in the paper*, it is not so much the actual Web documents (Web pages) that distinguish this perspective, but rather the context of usage (rapid and rich access, hypermedia linking, etc.). A crucial question for the field then is how we capitalize on this rich environment by designing optimal Web documents for learning purposes?

Design framework

An important issue that underlies design considerations is that of the directiveness of the learning materials. Under another guise, this issue of learner control has been with the field of computer-based instruction ever since its inception (Duchastel, 1994). On the one hand, some forms of instruction are very didactic and strongly guide the learner to the achievement of very specific learning outcomes. Formal instructional systems design (ISD) is the appropriate methodology for this approach (Gagné et al., 1988). Along the way are forms of instruction that create a context for guided exploration, often supported by more advanced intelligent CAI systems; here, the design methodology of choice may well be one derived from an artificial intelligence software design context (Duchastel, 1991). At the far end of the spectrum, we find fully autonomous learners guided more by their interests and personal learning goals (the situation of informal learning on the Web). It is in this realm that we see a strong need for a design methodology that is adapted to the particular style of interaction that the Web generally entails (Duchastel, 1990; Nielsen, 1995).

Learning context is central to adapting a design process to the web environment. In particular, the issue of level of learning is crucial. In traditional learning contexts, the level of difficulty is carefully controlled through adaptive means so that learning processes are optimized. Simplification and structural support (sometimes called intellectual scaffolding) are instructional means used to achieve this optimum level of learning. These, however, are generally not present in a typical web context, with the consequence that greater cognitive demands are put on the user-learner to make the best use possible of the information resources available. *The paper explores these cognitive demands and proposes a set of learning requirements based on the factors of interest, comprehension and retention.*

Personal context: Effort and interest (the effin factor)

The notion of an information-rich learning context greatly affects traditional forms of learning and instruction. An information-rich context means not only one in which a rich store of information exists, but more importantly, one in which there is easy access to that information. The reason the printed resource is so appealing today as a learning resource is quite simply that it is very readily available.

It is useful to ponder the notion of an ever-increasing information-rich learning milieu, for there is something at once fascinating and suspicious about the possibility that merely increasing information accessibility will lead to more and better learning.

Let's start by considering the current situation. Until recently (and even today), the typical student may not have had instant access to all desired information, yet there was a tremendous wealth of information available both in a home's encyclopedia and in the local public library. However, despite this vast wealth of information in the student's educational context, there was relatively little use made of it (certainly much less use than one might initially expect). Why is this ? The answer lies in the relationship that exists, at a psychological level, between personal effort and interest, what can be called the effin factor (an *eff*ort-to-*in*terest trade-off).

The effin factor concerns the appeal or lack of appeal of some activity or resource. It involves the effort a learner is willing to invest in pursuing an activity or in accessing a resource, in relation to the learner's current intrinsic interest in the topic. The effin factor deals with the motivation of the learner, but with intrinsic motivation only (Berline, 1960), one related largely to curiosity (because of its relation to knowledge, this is known as epistemic curiosity).

The effin factor is at the center of learning technology. It builds on our natural curiosity about the world around us, and on the potential of technology to make information easy to obtain. We are rather flighty in our curiosity and our attention keeps shifting as we encounter new and interesting things in our dealings with the world. It is important therefore that the satiation of our appetite for information occur at the time of need, otherwise there is a good chance that interest will be lost. Hence the importance of immediate availability of the needed information. Any delay increases the mental effort in keeping the interest alive.

The effin factor embodies a straightforward relationship: the more the effort needed to obtain the needed information, the more that topic must be of interest to the learner. If the required effort is great, only topics of very high interest will be pursued. On the other hand, if the required effort is minimal, then many topics become of interest. As web technology reduces the effort to access information, the potential for interest in many diverse topics grows for individuals. Natural curiosity has at last a chance to blossom.

The central idea developed here is that information is becoming readily accessible in a variety of compelling forms, and that this information-rich context can dramatically affect how students and people in general go about learning. From that notion follow implications for how we design instruction, that is for how we develop learning environments.

An initial design model

The design model presented here is framed by the need to interest the user on a continual basis. This can be thought of in terms of *provoking learning*, that is enabling it to spontaneously happen in the context of the user's interaction with web resources as this interaction takes place in a somewhat un-planned and perhaps even haphazard way.

The design model is couched within a standard software development life-cycle process based on a general systems approach. Thus, various elements of requirements specification conducted up-front and prototyping, evaluation, construction and support at the back-end help ensure a high level of quality of the products designed following the model.

The model's core specificity, however, revolves around 2 central design processes dealing respectively with information specification / representation and problem-solution tasking. *The paper presents the specifics of the model, including the core process activities based on structural knowledge determination, information mapping, multimedia representation, cognitive task analysis and interfacing.*

No design model is universal and it is important to situate the present model in terms of appropriateness. Specifically, how appropriate the model may be for different types of learning needs to be examined. *The paper discusses this issue in terms of appropriateness for procedural, declarative, and cultural learning. It also explores the issue from the older but still current Piagetian perspective of assimilation and accommodation phenomena in learning.*

Conclusion

The paper concludes with the view that the web context provides a novel setting for learning that is different in a fundamental way from its predecessor non-information-rich setting and that this new setting invites reflection and experimentation with new design processes that will facilitate learning from resources to be found in that chaotic and expanding setting.

References

Berline, D. (1960). *Conflict, Arousal, and Curiosity*. New York: McGraw Hill.

Duchastel, P. (1990) Examining Cognitive Processing in HyperMedia Usage. *Hypermedia*, 2, 221-233.

Duchastel, P. (1991) Towards Methodologies for Building Knowledge-Based Instructional Systems. *Instructional Science*, 20, 349-358.

Duchastel, P. (1994) Learning Environment Design. *Journal of Educational Technology Systems*, 22 (3), 225-233.

Gagné R., Briggs L. and Wager, W. (1988). *Principles of Instructional Design.* (3rd ed.) New York: Holt, Rinehart and Winston.

McAdams, M. (1995) Information Design and the New Media. *Interactions* (October 1995), 39-46.

Nielsen, J. (1995) *Hypertext and Hypermedia: The Internet and Beyond.* Boston: Academic Press Professional.

Extended abstract - the full paper is to be found at http://www.nova.edu/~duchaste/design.html

The Web is not a Distributed Multimedia System (Yet?)

Erik Duval, Henk Olivié
Departement Computerwetenschappen, Katholieke Universiteit Leuven
Celestijnenlaan 200 A, B-3001 Heverlee, Belgium
Email: {Erik.Duval,Olivie}@cs.kuleuven.ac.be

Introduction

The Web, although immensely popular, isn't really a distributed multimedia system, because the integration of different media is very shallow. Rather, the Web can be considered as a 'multiple media' system, where monomedium objects co-exist.

In order to turn the Web into a true multimedia system, support for spatial and temporal relationships between objects must be upgraded. In the present state of Web technology, these relationships are barely considered at all:

- In HTML, the <IMG> tag indicates that one object, a still image, must be included in another one. Limited indications with respect to spatial positioning of text around the image can also be included. A generalisation of <IMG> is the <OBJECT> tag, proposed in a WWW Consortium (http://www.w3.org) working draft, which also supports inclusion of applets, client-side image maps, etc.
- Conventional hyperlinks also define interrelationships between documents. But these are too limited: one either 'jumps' to another screen or starts a helper application. In both cases, one switches from one monomedium object to another.
- Co-ordination in time, i.e. synchronisation, is not considered at all in the WWW. In this context, it is useful to distinguish between continuous synchronisation (e.g. a video clip where the speaker's lips must remain synchronized with the sound) and event driven synchronisation where single points of different streams must coincide (e.g. a slide show with associated audio) [Ehley, Ilyas & Furht 1995].

'Web programming languages' like for instance Java make it possible to define more complex objects with associated behaviour. However, this additional facility comes at great cost: documents become closed (i.e. their internal structure is hidden from the Web client), and the Web becomes a platform for application delivery rather than a distributed multimedia system.

What is needed is a set of higher level constructs for defining the structure of Web documents. Since the beginning of 1996, we have started a project that investigates appropriate such constructs, to be superimposed on a flexible, set-based hypermedia data model [Duval, Olivié & Scherbakov 1995]. As explained in the next section, we are currently concentrating our efforts on time-dependent hyperlinks.

Time-Dependent Hyperlinks

We are investigating the concept of time-dependent hyperlinks in the context of our set-based hypermedia data model [Duval, Hendrikx & Olivié 1996], that integrates the more important features of earlier data models. In our model, objects belong to an arbitrary number of sets. The conventional link concept is replaced by a topology operator $\theta\, o,s \rightarrow x$: its parameters are a set s (the current context) and an object o (the current object, an element of s). The result is a set x, subset of s, with all the elements of s that are accessible when o is current. In order to model temporal characteristics of hyperlinks, the topology operator θ will be extended with time information, so that it becomes $\theta\, o,s,t \rightarrow x$, where t models the time that has elapsed since the display of o started.

As a concrete instantiation of the topology operator θ, we are currently working on an implementation where θ is defined by specifying begin and end times for an element to be accessible. This mechanism can be used to specify for instance that information on Bill Clinton is only to be made available when he appears in a particular video clip (say from 2.17 seconds after the start of that clip, until 8.35 seconds after that start). This approach can be generalized further, replacing the time information t by a more general condition c. At any moment during the play-out of o, the elements of x will then be accessible when c evaluates to true.

Another issue specific to time-dependent information in a hypermedia context, is what happens when a new element becomes current. A number of options exist for the previously current element: this element can

- remain active (play-out of a music clip can continue while the end user consults information about the composer);
- be suspended (an animation on a chemical reaction can be halted while a student browses information about a chemical element involved in the reaction; once browsing is finished, the animation can resume where it was suspended);
- disappear completely (a video clip on a news item becomes irrelevant when the user selects a clip on the next such item).

In order to deal with these options, the access operator that replaces conventional link following in our hypermedia data model must be extended.

Although the main focus of this project will be on conceptual modeling issues, we will design and implement a demonstrator to prove the viability of our concepts. For the server side, we will rely on the HOME hypermedia server we developed previously [Duval et al. 1995]. For client-side development, we will probably rely on Java. From our introduction, it should be clear that we consider this approach acceptable for prototyping purposes only, and that the relevant functionality, once mature, should be integrated in the Web in a more seamless fashion.

As far as related work in distributed hypermedia is concerned, spatio-temporal synchronisation has received relatively little attention. Besides the HyTime and MHEG standards, which define interchange formats for multimedia objects with synchronized components, the Amsterdam Hypermedia Model [Hardman, Bulterman & van Rossum 1994] and work on the Easy and Harmony viewers for HyperWave [Maurer 1996] are the more notable exceptions.

Acknowledgment

We gratefully acknowledge the financial support of the Belgian 'Nationaal Fonds voor Wetenschappelijk Onderzoek'.

References

[Duval, Hendrikx & Olivié 1996] E. Duval, K. Hendrikx and H. Olivié. Set-Based Hypermedia Modeling. MMM 96: 3rd International Conference on Multi-Media Modeling, November 1996.

[Duval et al. 1995] E. Duval, H. Olivié, P. O'Hanlon and D. G. Jameson. HOME: an Environment for Hypermedia Objects. Journal of Universal Computer Science, 1(5):265-287, May 1995. (http://www.iicm.tu-graz.ac.at/Cjucs_root)

[Duval, Olivié & Scherbakov 1995] E. Duval, H. Olivié and N. Scherbakov. Contained Hypermedia. Journal of Universal Computer Science, 1(10): 687-705, Oct. 1995. (http://hyperg.iicm.tu-graz.ac.at/Cjucs_root)

[Ehley, Ilyas & Furht 1995] L. Ehley, M. Ilyas and B. Furht. A Survey of Multimedia Synchronization Techniques. In B. Furht and M. Milenkovic. A Guided Tour of Multimedia Systems and Applications, pp. 230-256, 1995.

[Hardman, Bulterman & van Rossum 1994] L. Hardman, D. C. A. Bulterman and G. van Rossum. The Amsterdam Hypermedia Model - Adding Time and Context to the Dexter Model. CACM, 37(2): 50-62, Feb. 1994.

[Maurer 1996] H. Maurer. Hyper-Wave: The Second Generation Web Solution, 1996.

WEB-EDD: Electronic Document Delivery Via World Wide Web

Ed Dzierzak
Health Science Libraries
Marshall University, USA
dzierzak@muvms6.mu.wvnet.edu

Kent Hayes
Dept. of Finance & Administration
School of Medicine
Marshall University, USA
khayes@musom.marshall.edu

William Lucas
Dept. of Family & Community Health
Marshall University, USA
blucas@musom.marshall.edu

Until recently, document delivery was a rather slow process. In the past requests were sent and received through the US Postal service. In the medical field, the National Library of Medicine offers an electronic system for routing Interlibrary loan (ILL) requests - DOCLINE. OCLC is another such service used by public and academic libraries. Even with these electronic systems for request submission, the actual delivery still depended on the USPS. With Marshall's Rural Health Education Partnership program, providing information to our students and faculty at the rural sites is a major concern.

Fax transmission helped for items needed in a hurry, but quality, until recently, was not all that good. The Research Libraries Group stepped into the fray with *Ariel*, a method for transmitting the document images via the Internet. Finally, it was possible to get good quality documents quickly. This also had the side effect of making the ILL almost paperless through electronic transmission of the request and electronic transmission delivery of the document. Unfortunately, *Ariel* requires special software on the sending and receiving ends for the document transfer, and has remained largely a library to library transfer medium.

An *Ariel* file is a "standard multi-page TIFF file compressed to the Group IV fax standard and preceded by a GEDI (Group on Electronic Document Interchange) header." *Ariel* uses and "can import only standard single-page or multi-page TIFF files...". Therefore, to use *Ariel* files for transmission to the user would require a special viewer and FTP transfer of the document files. To provide this service to the rural sites requires a modem - not a particularly fast medium. A 150K file will take about 50 seconds to transfer with a 28,800 modem. The usual journal article requested from the HSL is about 8 pages. A little simple arithmetic shows that to transfer 8 pages takes about 400 seconds - 3.5 to 4 minutes. While this does not seem to be a lot of time, when a person is sitting at a computer waiting, it is perceived as something just short of eternity. The goal of this project is to provide a fast, easy-to-use system for making and receiving document requests.

This prototype, called ***Web-EDD***, uses PC-compatible hardware, and some custom software to scan the desired document, attach it to the Health Science Libraries' Web page, provide security for the transmission and automate the notification of the user. On the user's end, what is necessary is an Internet connection, PC-compatible hardware and a freely available viewer. A custom-written viewer is necessary to make the transfer process as quick and painless as possible.

Web-EDD is an integrated system designed to provide a friendly interface to access printer literature for any user with Internet access. An interface style similar to browsers currently used on the World Wide Web was chosen to minimize or eliminate the additional training of both the end users and library staff members. The process follows the normal method of obtaining a copy of a document. The user using a Web Access Interface call *EDD-View* initiates the process. *EDD-View* is an integrated program which provides users the means to submit a request for a document, to retrieve the scanned image of the document and to view the facsimile of the document. After gaining login permission to a *Web-EDD* system, the user initiates a request by completing a Web-based form which is submitted to a CGI program running on a Web Server (*EDD-CGI*). All requests for documents are managed through a work queue in *EDD-Cap* by the system administrator. Then documents are scanned with *EDD-Cap* and made available to the user for viewing. *ARIEL* files received from other institutions can be imported into the *Web-EDD* system through *EDD-Cap*. This allows *Web-EDD* to be integrated with an existing ***ARIEL*** system. Once a request has been completed, *EDD-Cap* automatically notifies the user by E-mail that the request has been filled. With a facsimile of the document now available, the user accesses his/her private listing of article requests through *EDD-View* for retrieval of the article. *EDD-View* retrieves and displays the selected article. by default, each page of a multi-page document is retrieved independently. This permits the user with slower connections to immediately view pages while the remaining pages are retrieved in the background. The user can access any page as soon as it has been made available within the viewer. The option to retrieve only selected pages of a document for viewing is permitted.

There are some considerations that have gone into this document delivery system. Security is maintained by making the documents available to the requestor only. This is done by having the requestor log into the document delivery system. The requestor could be anywhere, using any Internet-connected computer and yet still be able to get their documents. This aspect also allows for proper billing information to be collected. It does not matter whether the requestor is using their own computer or is across the country using a guest account. The requestor must be registered and authorized by the home system before requests are accepted and delivery done.

Copyright is a serious consideration in any form of document delivery. To stay within the copyright guidelines, the images will not be permanently stored on the document server. Once the document is received by the requestor, it will be deleted. This is similar to what occurs in *Ariel*. The document is scanned at the owning library, transmitted to the requesting library, printed by the requesting library and then deleted. This is essentially the same as the photocopy and fax process. Not only would long-term storage of the images be illegal without permission of the copyright holder and payment of the associated royalty, a practical consideration is the long-term storage of these images would require a large amount of disk space.

As with any project, many questions come to mind. The decision to avoid OCR was significant. At this point, scanning the document and using OCR to change the text of the document into actual text is not practical. Most of the documents involved have some type of image (pathological specimens, etc.), graphs or tables. OCR cannot handle these sections and they would have to be transmitted as images in any eventIf the document is not placed perfectly on the scanner, it is possible for letters or numbers to be changed. This would be unacceptable for documents to be used for patient care. Research by Le, Thoma and Weschler at the National Library of Medicine is aimed at recognizing and correcting the skew angle of scanned documents.

Why not just use fax to transmit the documents? Fax quality is much better than it was 5 years ago. Fax is a good method for transmission of documents. WEB-EDD gives the added capability of viewing the documents before printing. Attaching the document to e-mail is another possibility. However, current projects such as the National Library of Medicine's DOCView, the Australian Development of Library Infrastructure, Dale and others use e-mail require MIME-compliant e-mail to accomplish the task.

Bibliography

Blinco, K. (1995) Some Recent Initiatives in Electronic Document Delivery. Online/Ondisk 95.

Group on Electronic Document Interchange (1995) Electronic Document Delivery: Towards Further Standardization of Internation Interchange. V. 2.1, http://www.gu.edu.au/alib/iii/docdel/gediv21.htm

Le, D. X., Thoma, G. R. & Weschler, H. (1994) Automated Page Orientation and Skew Angle Detection for Binary Document Images. Pattern Recognition, 27 (10), 1325-1344.

Walker, F. L., & Thoma, G. (1994). Access to Document Images over the Internet. Online Library Systems, New York, NY, 185-197.

Walker, G. L. & Thoma, G. R. (1995) DocView: Providing Access to Printed Literature through the Internet. Online Library Systems, Medford, NJ, 165-173.

Walker, G. L. & Thoma, G. R. Internet Document Access and Delivery. Information Today

1996, 107-116.

Customizing Web-Based Course Delivery in WEST with Adaptive Navigation Support

John Eklund
School of Teacher Education, Faculty of Education
The University of Technology, Sydney
j.eklund@UTS.edu.au

James Sawers
Educational Multimedia Unit
The University of Technology, Sydney
james.sawers@UTS.edu.au

1. Flexible Delivery of Course Materials

New technologies to implement on-line education systems are a key component in the flexible delivery of courses. A key problem with on-line learning identified in recent literature [Pennell 1996, [HREF2]; Jones 1996a, (HREF3)] has been to provide quality interaction with students through individualised instruction, customised course material and a personalised dialogue with learners. Recognising this, our aim was to implement a system which provides immediate feedback to students and attempts to address the problems experienced by learners in navigating hypermedia courseware. This paper describes a flexible delivery initiative currently underway at The University of Technology, Sydney (UTS) to introduce partial on-line delivery of course materials to Educational Computing students using the Web-based authoring tool WEST [HREF1]. We have used the simple and intuitive interface of WEST to deliver courses on the Web, with a modification which provides visual annotation of links [Brusilovski & Pesin 1995, (HREF6)] in the courseware to provide navigation support. On the basis of the student's results in on-line testing modules, the system provides individual navigation support to guide the learner through the hyperspace. The courseware presently consists of eight subsections each of which has a set of outcomes, a prior knowledge test, pages of instruction, a set of exercises to be completed for the face-to-face sessions, and a topic test. The system is currently adaptive [HREF5] only in the simplest sense, but is specifically being used as the basis for an empirical study which aims to determine whether link annotation [Brusilovsky, 1996] to provide individual navigation support improves learning outcomes for students.

2. Modifying the WEST System for Navigation Support

The architecture we are implementing places WEST within a frameset in which the main frame contains the course pages, and the navigation buttons of WEST, just as the WEST interface usually appears. The course pages include hyperlinks to relevant materials on the Web, so that students may investigate resources already on the Web related to the topic. These hypertext links are targeted to a "blank window" so that students may browse outside the courseware and easily find their way back. In the bottom frame is a clickable overview map that is generated for each student which shows the hypermedia nodes and annotates the links as visited, unvisited, current and suggested. After further development, we aim to generate this link set automatically, so that as pages of courseware are added, the overview map will update itself. It will also show what page has been learned, and what pages have yet to be mastered. This requires further work in analysing the content of the pages and the individual questions in the online tests. In this way, the link set becomes a "glass box" user-model as well as a navigation aid: the student can see what progress he/she is making and what the system understands about that progress.

A somewhat similar (and fully implemented) system is ELM-ART [Brusilovsky, Schwarz and Weber 1996], a critical difference between the ELM-ART system and what we are developing using WEST is that in ELM-ART, once a student has visited a node, the system assumes that the student has read and understood the material at that node. In this way ELM-ART assumes that students are purposeful and deliberate about their movements through the hyperspace, it does not account for behaviours such as looking ahead at material before reading it, or reading material and not understanding it.

3. Evaluation of Link Annotation for Navigation Support

With the system under development, we are conducting an empirical study to see whether the link annotation is effective in increasing learning outcomes. This is being done by disabling the adaptive navigation support for a random half of the students for the first part of the course, then enabling it for the second, and vice-versa. We are using audit trails and the results of the online tests for both the formative and summative evaluation. The presentation of this paper at WebNet96 [HREF4] will focus on the development of the system and report early results from the empirical study.

4. References

[Brusilovsky, Schwarz and Weber 1996] Brusilovski P, Schwartz E & Weber G [1996] ELM_ART: An intelligent tutoring system on the World Wide Web. In Frasson C, Gauthier G & Lesgold A [eds.] Intelligent tutoring systems. Proceedings of the third international conference, ITS'96. pp 261-269. Springer, Berlin.

[Brusilovsky 1996] Brusilovsky P [1996, in press] 'Methods and techniques of adaptive hypermedia'. User Models and UserAdapted Interaction 6

[Jones 1996b] Jones D [1996b] "Computing by Distance Education: Problems and Solutions", Proceedings of the ACM SIGCSE/SIGCUE Conference on Integrating Technology into ComputerScience Education, Barcelona

5. Hypertext References

HREF1: WEST home Page

URL:http://www.west.ie

HREF2: Pennell R [1996] Managing Online Learning. Paper presented at Ausweb96: The Second Australian World Wide Web Conference

URL: http://www.scu.edu.au/ausweb96/educn/pennell/paper.html

HREF3: Jones D [1996a] Solving some problems of University Education: A Case Study. Paper presented at Ausweb96: The Second Australian World Wide Web Conference

URL: http://www.scu.edu.au/ausweb96/educn/jones/paper.html

HREF4: Home Page of WebNet'96

URL: http://www.websoc.at/0x811b9911_0x00033a17;sk=800B0278

HREF5: The Adaptive Hypertext and Hypermedia Home Page.

URL:http://www.edfac.usyd.edu.au/projects/ah

HREF6: Brusilovsky P & Pesin L [1995] Visual annotation of links in adaptive hypermedia. Procedings of Chi95

URL: http://www.acm.org/sigchi/chi95/Electronic/documnts/shortppr/plb_bdy.htm

An Innovative Teaching Technique Using the Web to Deliver Case Study Instruction

Dr. Ernest B. Ezell, Jr.
Department of Visual Communication and Technology Education
Bowling Green State University (Ohio) USA
ezell@bgnet.bgsu.edu

Dr. Diane K. Frey
Department of Family and Consumer Sciences
Bowling Green State University (Ohio) USA
dfrey@bgnet.bgsu.edu

Background

The World Wide Web is a unique technological resource that allows information to be presented in a dynamic interactive format. This includes presentation of text as a document; words as links; tables that summarize; and graphics that show a virtual photographic gallery. Dynamic formats exist for individual feedback and assessment of student learning and analytical development. Caffarella (1993) addresses the issue of self directed instruction as it relates to adult learning. In a digression from traditional learning theories, he presents evidence "that many adults can and do learn primarily through their own initiative". One of the challenges of self directed instruction is in making the transition from teacher control to learner control; except in terms of the Web, the facilitator is the Web and its multitude of hyperlinked resources.

Does this sound vaguely familiar? Of course it does! This may be that new catch phrase that we are all hearing about, called "distance learning". Using the Web to scan the expansive amount of materials is a form of distance learning or distance education. Distance learning is any form of teaching and learning in which teacher and learner are not in the same place at the same time; with information technology their likely connector. Over the Internet, individuals are presented multiple opportunities for accessing information from around the world.

Students can and do learn independently, in groups, from reading and projects, and in many other ways. Literature on learning theory and practice encourage instructors to facilitate a variety of student-centered activities. The self directed student, as do other students, must be able to assess, think and communicate. The thinking student must develop and process information and apply the information to real situations. This is a form of problem solving.

As the technology helps us to break away from the hegemony of an in-person, here-or-nowhere-else view of learning, new possibilities open up. Distance education requires thoughtful attention to pedagogy and to the settings in which learning can occur. The sheer power and rapid improvement of telecommunications options make it imperative that we identify how to best use the combinations of face-to-face, independent, and "distant" learning. As we do so, we'll find insights that help us rethink what we do in any form of education.

Specific Problem

As issues within the apparel industry continue to become increasingly global and complex, traditional problem-solving paradigms may no longer be optimal. Students who take jobs in apparel related fields may be required to effectively deal with challenges that require complex thinking and sophisticated problem-solving skills. However, there is concern that college students are not being assisted in developing specific thinking skills that will help them contribute effectively to business. Thus, there has been an increasing emphasis on meeting this challenge through learning strategies to assist the student in thinking critically.

Solution

Based on the needs addressed above, the investigators elected to utilize the Web as a delivery mechanism for presenting case study instruction to college students (Frey, Eckman & Ezell, 1996). The objectives of the proposed project were to: 1) develop case studies that are relevant to current issues; 2) develop an interactive Web based approach to deliver the case studies and enhance problem-solving skills. (The intent of the program was to build around the case study method of instruction; this serves as a medium for encouraging students to consider a variety of perspectives and possible solutions and to develop a rationale for a decision.)

The unique aspect of this project was to develop an interactive Web based program featuring case studies that presented issues addressing the ever changing nature of international trade issues (i.e. NAFTA) promoting free trade among nations. This poses economic challenges for the domestic manufacturers and retailers in developing strategies for growth in a competitive world market. The Web has the potential for facilitating the application (i.e., case studies) of problem-solving skills while users gather current information, identify issues, propose solutions, receive feedback, and gain experience through solving problems. Interactive computer systems that encourage communication among learners and instructors may be more successful at teaching problem-solving skills than when students work alone or toward individual goals.

The primary function of the Web instruction was to assist students in developing problem-solving skills and to receive feedback through dialogue with other students and the faculty supervisors. The program would ask the student to identify critical issues related to a specific problem and propose possible solutions to the problem. Students have several links available in researching the issue: government agencies, professional associations, and industry. Students will also be directed to additional instructor developed resources for obtaining relevant information about the case.

It is possible also to simulate different roles of a business owner (manufacturing or retail), economic or marketing consultant, a member of a merchandising/product development team, private citizen, representative for a resident buying office for retail stores, or representative in government. Each student will write a report summarizing and interpreting the information that they have gathered. This information is then collected through a Web form and the file is then transferred by e-mail. A collaborative file called "Recommendations" is prepared and read by each student. Reactions will become the basis for an effective "discussion" through the internet. Finally a collaborative decision will be reached. The use of quality management procedures will be incorporated throughout the case studies and the problem-solving approaches to issues can be applied to any case study appropriate for industry.

The effectiveness of these instructional strategies was assessed through the achievement of the students. The program was evaluated based on 12 criteria adapted from a refereed questionnaire (Barker & King, 1993). Students evaluated the program's ability to engage the user's interest and ease of use from choosing, entering, or considering possible solutions. Program characteristics were also evaluated from the quality of interaction, quality of content, and noting any strengths and weaknesses. Both quantitative and qualitative evaluation procedures were conducted.

Summary

The case study project involved developing an interactive Web based program which allows associative structures of dynamic information and decision-making. Educators have long believed that active learning brings about better understanding of key lesson concepts. Furthermore, it is important that instructional designers recognize that different people use learning opportunities in different ways. The World Wide Web is a unique technological resource that allows information to be presented in a dynamic interactive format. The developers have taken advantage of this resource and organized a small portion of the available content to present case study instruction in a very original format.

References

Baker, P. & King, T. (1993). Evaluating Interactive Multimedia Courses- A Methodology. Computers in Education, 21 (4), 307-319.

Caffarella, R. (Spring 1993). Self-Directed Learning. In S.B. Merriam (Ed.). An Update on Adult Learning Theory (pp 25-35). San Francisco: Jossey Bass.

Frey, D.K., Eckman, M. & Ezell, E.B.(1996). The Impact of NAFTA Upon the Apparel Industry, [On-line]. Available: http://ernie.bgsu.edu/~ezell/Case_Studies/intro.htm

Establishing Distributed Learning Communities (DLCs) using the Internet

Kevin C. Facemyer
Virtual Professional Development School
College of Education
Washington State University, USA
KCFacemyer@wsu.edu

Gary Brown
Center for Teaching and Learning
Washington State University, USA.
browng@mail.wsu.edu

Nils Peterson
Virtual Professional Development School
College of Education
Washington State University, USA.
Nils_Peterson@wsu.edu

This short paper highlights recent developments at Washington State University (WSU) in establishing and building Distributed Learning Communities (DLCs) utilizing the Internet; the different styles of Internet use: Presentation, Interactive or Generative and how each contributes to generation of communities; and an analysis of new virtual leadership strategies for the Internet.

In March 1995 the WSU Virtual Professional Development School (vPDS) invited the world's P-12 students and teachers to participate in worldís first Virtual Science and Mathematics Fair (vSMF) hosted on the Internet (This yearís Science Forum is at: http://www.vpds.wsu.edu/sciforum). Students were asked to formulate a question (pose a hypothesis) and research it. "Does music affect mice's ability to learn a maze?", or "Is there a difference in confections created with generic versus name-brand ingredients?" were the types of question students asked and then answered. The students investigated formally constructed hypotheses, not by library or reference methodologies, but by running mice through mazes or baking treats and having human volunteers taste test and indicate preferences. Some of these investigations were done collaborativly at a distance using the Internet to share data and discusisions. These results were presented in a virtual poster session. Students and the public were invited to review and discuss the results with the student scientists. The resulting DLC allowed students to understand that science is conducted not by individuals or even local teams, but by teams of varying backgrounds, philosophies and theoretical frameworks distributed around the world.

The vSMF participantsí use of the Internet to gather and share data, to make presentations and have discussions has brought young P-12 student-scientists very close to participating in the critical features of science as a global enterprise. The features of hypothesizing, investigating, presenting and defending their work were all enhanced by the use of the Internet to establish the "virtual conference center.î This empirical research in a collaborative setting, some using distant research collaborations, is very similar to the research and presentation methodologies of professional scientists in several different fields.

For instance, the Princeton Plasma Physics Laboratory (PPPL), has begun to broadcast over the Internet sound and video from the control room of its Tokamak Fusion Test Reactor. These results and ensuing discussions have participants from around the globe. (Butler, 1996) Several other types of distant experiment and control, such as high voltage electron microscopy, radioastronomy and automated protein modeling are made possible to remote users by the Internet. That children have discovered similar types of protocols based on the Internet's ability to aid in communication and control, suggests that these are more advanced modes of investigation and collaboration than were previously possible without the Internet. The DLCs arose in response to need rather than edict, and for a purpose similar to professional scientists (*generation* of knolwedge). This suggests that DLCs will become an increasingly common work environment as these children grow up and and become professional users ot the Internet.

As students mature and continue to use the Internet in advanced ways, they will use the Internet in a varieties of ways. We describe the distribution of the Internetís use. Three points of use along this continuum are: Presentation, Interaction, and Generation--or PIG.

The Presentational use of the internet includes perusing enhanced lectures or using the World Wide Web as library enhancement or as research resource. Presentational uses of technology reflect traditional pedagogies in which the teacher lectures, students listen; the teacher defines the research task, students report their findings. The nature of the PIG as a continuum becomes increasingly clear when a student asks an instructor posting lectures on the Internet to enchance a lecture or to clarify a point--pushing the presentation up the PIG toward Interaction.

Interactive use of the web includes prepared web-based modules that branch from site to site or within a site. The interactivity increases with the complexity of the branching, from Skinnerian right-wrong modules to more complicated Crowderian branches. At this point in time, most Interactive uses of the web extend only marginally away from the Presentational end of the continuum and are comprised primarily of text material or linked texts. As we move further up the PIG continuum, Interactive use also includes asynchronous threaded discussions on the web that may exhibit increased collaboration. Higher yet on the Interactive continuum, Internet participants interact, with not just text but with each other. Very soon after this type of Interaction, participants naturaly begin to Generate ideas that go beyond the contents and intents of the original hypertexts.

The Generative use of the Internet describes pedagogies that put multimedia web tools into the hands of students. This allows and encourages them to inquire and generate based on their interaction with the hypertexts and each other. Most of these approaches at WSU require students to collaborate with each other and with faculty to create web pages or multimedia modules. A number of independent studies at WSU have begun to establish that, in general, as pedagogies move up the PIG continuum into the Generative realm, exciting things happen. Students participate more, encounter more diverse points of view, are more likely to seek and find expert as well as peer resources, conceptualize material more completely, and, essentially, learn more. At the same time, the more pedagogies move toward the Generative, the more complicated they become, the more a change in tradtion becomes necessary. Generative pedagogies challenge traditional educational systems, institutions and organizations.

This necessitates a new virtual leadership strategy. We understand the overall vSMF leadership strategy as an example of the temporary systems model (Miles, 1964, Goodman & Goodman, 1976) ---students, adults and vSMF staff and administrators came together in ad hoc ways to solve specific problems related to participation in the vSMF. Many of these relations arose spontaneously among participants as they individually worked on technical or scientific problems.

An important dimension of the leadership was that it could not be coercive. Participants could leave the vSMF site at the click of a browser. The administrators adapted techniques known from other leadership settings, being visible in the hallways (web presence), leading by example (model posters), and talking up the process (listserv). This leadership mimics Sengeís (1996) Networker leader who is followed because of their compelling vision. In an environment of egalitarian knowledge distribution; where few sanctions or rewards exist aside from the gratification of participating in interesting, rewarding and enriching experiences; this combination of techniques becomes a model virtual leadership strategy.

Internet's role in supporting DLCs, its spectrum of uses, and the new styles of leadership is the basis for this paper. Distributed Learning Communities (DLCs) will play an predominant role in the future of education based on Generative pedagogies and the Internet. Understanding how to build and nurture DLCs with new virtual leadership strategies will be critical to administering successful educational environments and reforming educational institutions.

References

Butler, D. (1996), Research and the Internet connection. *Nature 380*(6573) 377-78.

Miles, M. B. (1964). *Innovations in education*. New York: Teachers College Press.

Goodman, R., & Goodman, L. (1976). Some management issues in temporary systems: A study of professional development and manpower--The theater case. *Administrative Science Quarterly, 21*(3), 494-501.

Senge, P. (1996). *Leading learning organizations: The bold, the powerful and the invisible.* In the leader of the Future, F. Hesselbein, M. Goldsmite, and R. Beckhard. San Franscisco: Jossey Bass.

Introduction and Description of the Interactive Image Format (.iif) for the World Wide Web

Jeffrey Ferguson
The Weather Underground
Dept. of Atmospheric, Oceanic & Space Sciences
University of Michigan, USA
jaf@umich.edu

The World Wide Web has proved itself as a viable transport of interactive documents. Our community continues to create artistic, useful, functionality which piggyback's the enormous strides of raw computational power. The World Wide Web hosts interactive experiences which evoke emotions, and it provides a communication medium introduced this decade. It is necessary to provide an interface which captures the inovative experiecne the WWW has to offer. 'Page' is a notion which attracted large numbers of users and creators. The Web has redefined what can be displayed on a 'page of text'. The interactive Image Format assumes its underlying foundation to be an image. Text (and other media) are placed 'within' the image and are revealed as a user interacts with the image (mouseOver, mouseClick). This technology is utilized on interactive CD-ROM, however Interactive Image Format (.iif) is a language to define this functionality on multiple WWW platforms.

Introduced here is a language used for constructing interactive images. The Weather Underground Inc., of Ann Arbor MI, publishes real-time (less than 1 hour old) interactive meteorlogical imagery to the WWW community. Unique with this meteorological imagery is the ability for users to receive immediate, server supplied, relevant feedback in relation to the cursor. In other words, when one passes over a city, the current conditions are displayed, and a mouseClick retrieves that city's forecast. A client retrieves an Interactive Image Format (.iif) file which supplies links to imagery and text to create the interactive image.

Many WWW browsers have similar functionality built into the interface, however the displayed text is an URL, which can often be unintelligible. IIF defines language constructs which provide the functionality necessary to generate actions on given regions, including the displaying of.customized text based on cursor location. IIF is the language which defines a set of regions with associated reactions to user actions. IIF files (ascii text files) are interpretted at run time. The IIF language consists of things, actions, and reactions. Typical things are: Recatangle, Circle, Polygon, RGBColor. Typical actions are warm (mouseOver), hot (mouseClick), and timer. Typical reactions are Show, Hide, GetURL, Front, and Flash. This lexically small language provides a great deal of interactivity for a successful user interface. IIF can animate icons, rolling-highlight of choice-lists, timed screen refreshes, pop-up text/menus. IIF extends the interface of the isMap by providing immediate, relavant feedback.

The IIF (text file) is interpretted by a Java applet, and is referenced as a .HTML document. The text file is small and travels quickly across the network. Java insures the applet is cached on the client, and only the .iif text files are transmitted. IIF's can be built nested within each other, and thus providing a method for graphically traversing the WWW. Images are linked to images, and any explanatory text is hidden from the user, keeping a tidy interface. When a piece of the interface is queried (mouseOver) by the user, relavant information (text, animation, sound) is revealed. This is in contrast to an interface which assumes a textual environment with images as perifery. IIF's ability to interpret warm (mouseOver) actions provides an interface functionality currently missing on the World Wide Web.

Anatomy of IIF

IIF's consist of a text file in the following generic format:

```
named-thing1(thing-type (parameters & location)
        action1 (
        reaction)

    action2 (

        reaction))

named-thing2(thing-type (parameters & location)
    action1 (

        reaction)

    action2 (

        reaction))
```

Why would someone use .iif?

The World Wide Web has created an interface upon the internet through the use of a mark-up language, HTML. The expedient growth of the WWW was the fact that HTML was a sophisticated language which was straight forward, so as to allow non-programmers the opportunity to create interesting Web sites. An HTML browser with an embedded Java interpreter provides the WWW community to explore interactive interfaces, however, ramping up to writing Java code is unattainable for the majority of WebSite Developers. It is important to consider the reasons for the explosive growth we have witnessed in the past few years. IIF has abstracted many interactive interface components into the IIF language. As HTML, new language constructs can be incorporated into the language syntax. For end-users to take advantage of language enhancements, the Java applet that is shipped with the HTML page will contain the upgraded syntax. As a WebMaster, an

The Web in Distance Education A New Technology for Some Old Problems

Richard Ferguson
The Swedish U. of Agricultural Sciences
Dept. of Economics, Sweden
e-mail: richard.ferguson@ekon.slu.se

Abstract: In this short paper, discussion of the implications of using the Web as a media in distance education is focused mainly on the affects on dialog. The author, however, wishes to stress that the choice of technology and media in distance education is multifaceted. It is important that instructors base their decisions on balanced pedagogic merit and not on single issues, personal interests, or fads of the day. For a more expansive and general discussion of choosing a media, readers are referred to Sören Berglund (1993) and/or Sue Spencer (1986).All successful distance education has to overcome the problems associated with the instructor and the student being separated by geographical distance. This distance complicates building a trustful and constructive student-teacher relationship. Additionally, students may loose enthusiasm in the course because of the lack of social contact as well as because of short-comings in the increased need for self-discipline. And not least, tutoring may be inhibited by the delay in feedback to questions and problems. While the trend in delivery technologies in distance education has been towards faster, more direct means of communication-having progressed from letter writing to telephone and fax to computer-simulating classroom dialog is difficult. In this light, how can and should the Web be seen and used?John Bååth (1996; 1994) distinguishes between the material (M) and the dialog (D) components in designing a course for distance education. Similarly, Holmberg (1989) defines the two constituent elements of distance education as the presentation of learning matter and the interaction between students and instructors (including the supporting organization). Both educators say that the dialog or interaction is important for motivating students, supporting learning, activating the use of new knowledge, and facilitating assessment of students' progress. The value of the Web for presenting the material is fairly obvious, though there are some comments to be made, but what are the consequences of a Web-based distance course for the dialog between students and instructors?Bååth emphasizes the importance of preparing a study guide written in a personal, conversive style to help students through the course literature. The study guide is not only a supplement to other written material (M-component), but is also an important link in establishing a "dialog" with the students (D-component). Through the written words, the instructor can reveal a bit of him- or herself to help build a rapport for a better student-teacher relationship; suggestions can be made to make reading and understanding the course material easier; questions can be posed to stimulate reflection; and assignments can be given, providing an opportunity for two-way communication. The instructor should not only see the study guide as a presentation of information but, through the tone set by the style, as a means of opening a dialog.A paper-based study guide must be used together with some means of communication-traditionally, the postal service-if a dialog is to be conducted. In this way, an instructor and a student or students can have a non-contiguous dialog, exchanging information and ideas, questions, assignments, feedback, etc. The advantages of postal-based communication is the relative accessibility and ease of use, but these advantages are in part off-set by the time delay between the messages from the participants. The increased use of telephone and fax in distance education has greatly decreased the time lag in distance dialog. But the telephone, though usually quite accessible, requires all parties to be present at the same time-thus detracting from one of the major advantages of distance education-and facsimile makes sacrifices in accessibility and ease of use. On first consideration, the Web may be seen as merely the next step after the fax. It provides speedier transmission and text without the paper, but with yet more sacrifice in access and ease of use. This technology dilemma is worth noting: Even though new technology addresses short-comings in current practices, the "higher" the technology, the more limited is access and the more knowledge is required for use.Integrating the use of the Web into distance education, however, has more to offer than simply faster, more convenient communication for those who have access. By preparing a study guide that takes advantage of what the Web has to offer,

a more interactive experience can be offered to students, stimulating their motivation, curiosity, learning, and dialog. In the remainder of this paper, a number of specific ideas for utilizing the Web to stimulate dialog in distance education will be discussed.

- Basic HTML formatting can be used to create a study guide that is easy and pleasant to use. Proper use of headings and text styles make pages that are easier to read. Authors should keep in mind downloading time, remembering that students may be connecting from home via a modem. Instructors should design pages that take advantage of HTML's navigation capabilities: Include a navigable table of contents; New terms can be linked to definitions; Use a standard navigation bar at the bottom of each page to get students to the next page, back to the table of contents, to assignments, to help, or to any other frequented page.
- The ability to easily include graphics in text should be exploited. Again, instructors should keep in mind downloading time, but useful graphics (and even the occasional decoration) can enliven and clarify text. Charts, illustrations, and pictures can be utilized to help students to understand and to give them a visual reference for discussion. Also, including a picture of the instructors on the introduction page, and elsewhere perhaps even of the students, can give a face behind the voices. This simple effort may make some students feel more comfortable and willing to voice their opinions in discussions.
- All pages should have a link to the instructors' e-mail. People are becoming increasingly more comfortable communicating via e-mail. And, as far as speed and convenience is concerned, well functioning e-mail is hard to surpass. Of course, successful dialog via e-mail depends on the instructors' rapid response to students' questions and comments. Instructors might want to suggest taking certain discussions to a more open forum, where more students could participate. While this can be done by forwarding e-mail messages, conferences are often more practical (to be discussed below). E-mail is also an excellent way for students to send in assignments, though clear instructions should be given concerning file type and format.
- CGI forms can contribute greatly to a more interactive experience. Questions that students can answer "on the spot" can be integrated into the text, allowing students to "comment" directly during their learning. This technique can be used to allow the student to give feedback to the instructor, to reinforce learning, and to check comprehension. By asking the student to provide input during the learning experience, there is a greater chance of achieving active learning, where the student reflects on the material being introduced. By directing responses to a database, instructors can easily monitor students' progress and provide feedback.
- Setting up a special conference or chat group where students can freely discuss course issues, as well as be assigned discussion topics, can encourage students to use their newly gained knowledge and learn from each other. Experience has show that an initial assignment, requiring participation, greatly helps to get students freely participating in an electronic conference.

These few simple suggestions show that the Web can be a useful medium in distance education. While the problems with using the Web in distance education-most significantly, student access to the internet and the major investment required in learning advanced HTML authoring-have hardly been addressed, the potential the Web offers for creating an active learning situation make it worth serious consideration for distance education.

References

[Berglund 1993] Berglund, S. (1993). Teknik och media i distansundervisningen, rapport nr 3, Distansutbildning i utveckling (Technologies and media in distance education, report nr 3, Distance Education in Development). Umeå, Sweden: Umeå Universitet.

[Bååth 1994] Bååth, J. (1994). Distansutbildnings grunder (The basics of distance education). Saltsjöbaden, Sweden: DU Konsult.

[Bååth 1996] Bååth, J. (1996). Handbok för distans utbildare (A handbook for distance educators). Stockholm: Utbildingsförlaget Brevskolan.

[Holmberg 1989] Holmberg, B. (1989). Theory and practice of distance education. London: Routledge.

[Spencer 1986] Spencer, S. (1986). A model for selecting distance education delivery systems. Continuum, 50, 143-152.

AHP: Academic Home Page

Timothy Foster, M.A.
Office of Instructional Development
Northern Arizona University, USA
Timothy.Foster@nau.edu

Marilyn Fritz, M.A. Ed.
Office of Instructional Development
Northern Arizona University, USA
Marilyn.Fritz@nau.edu

Introduction

The Academic Home Page project at Northern Arizona University was implemented to provide faculty with an ability to efficiently, and with relative ease, develop an effective interactive academic home page. Key features of the template produced by the development team include a clean user interface design, information flow design, balanced graphics, and instructional strategies. The visual appearance of the Academic Home Page has been modified slightly and produced in several versions to appeal to the personal preferences of faculty. An automatic HTML page creation process using forms for text input was considered, but it was decided that faculty developing home pages using this method would not have the knowledge and skills necessary to maintain their home pages.

The templates link to the Office of Instructional Development home page Instructional Toolkit, a critical component in assisting faculty in technical maintenance, as well as improving the instructional efficacy of this medium. The Instructional Tool Kit provides useful information about how to connect to the campus computer services such as PPP logins, listservs, news groups, and e-mail. This kit also contains innovative instructional strategies for use with the instructors weekly activities.

Faculty interest, as well as a call from the western governors' to create a virtual university, has also provided a new arena for world wide web development. Enhancing and teaching courses on the internet is now in the forefront. However, quality of delivery is a growing concern. The Academic Home Page and Instructional Tool Kit help assure a minimal level of quality that is superior to many instructional home pages currently on the web. The Tool Kit also provides innovative strategies for incorporating proven traditional instructional strategies into the web environment.

Academic Home Page Overview

The Office of Instructional Development (OID) has a mandate to facilitate faculty implementation of new technologies for instruction. OID strives to provide solutions to problems faculty encounter when they integrate technology into instruction. Although the internet is nothing new, the evolving world wide web environment offers unique possibilities for instruction.

The world wide web environment demands a new approach to designing and publishing information to take advantage of its inherent strengths. The ability to incorporate graphics and link to internal and external sources provides a powerful medium for creating more dynamic documents.

World wide web pages have been used predominately for information distribution only. An example of this use can be seen on numerous web pages. These web pages are not only text intensive, but they appear to be designed like most word processed documents. This graphic web environment supports much more interactivity than traditional word processed documents. The HTML language supports many new and exciting features such as animation, and even 3-D environments such as JAVA. One of the more powerful interactive aspects of designing an HTML document can be found in using links to other documents. Internal linking within the web page is also an important consideration in order to improve navigation within the document. Another advantage of using linking is providing non-linear access to a variety of information in a seamless fashion.

Unfortunately, regardless of the improving and user friendly internet browsing software and applications, many people have not had the time nor the inclination to learn a new language (in this case HTML). Locally, at Northern Arizona University a new UNIX machine was purchased and integrated into the campus computing environment. This machine provides faculty and staff with the opportunity to make their personal home pages public. Regardless of access to this UNIX server, a small percentage of N.A.U.'s faculty and staff have published their personal home pages. This presented our office with a challenge: that of providing solutions to typical problems and barriers faced by faculty using new instructional technologies. These problems and barriers include lack of time to develop a home page, and skill limitations in areas such as HTML tagging, information design, graphics design, creating an efficient home page, as well as designing pedagogical strategies for web based instruction. OID has brought together a team that incorporates all of these skills to create an exceptional template. This template helps faculty not only respond to their perceived needs, but also challenges them to think more creatively about using this new environment for instruction.

As the project evolved we designed five different templates specifically for academic function. There are minor variations in graphic appearance, for example in button design and color selection. We also collaborated closely with the campus computing technology services department to produce the templates. This aided our design process by ensuring the templates were efficient in terms of graphic download time and their appearance. The size and quality of the images used in these templates were also considered in order to provide end users with functional and attractive home pages.

Conclusion

Learning to edit and modify the Academic Home Page template provides faculty with opportunities to learn HTML and at the same time create a fully functional site for their students' use. The OID team combined skills in many different areas to create an effective set of templates for academic use. This collaborative effort resulted in a much richer set of templates.

Helping Teachers Use Anti-Virus Sites on the World Wide Web

Lynn Fox, Ph.D.

World Wide Web access for students and educators demonstrate how today's technological tools bring new roles and obligations to school administrators and classroom teachers. By necessity, their proficiencies will expand from the level of computer "newbies" (novices who have minimal computer skills) to the level of responsible computer users who know how to reduce the chances of viruses infecting their software programs and destroying their student's homework.

To start surfing wonderful web sites and go gliding past viral contaminated waves, school administrators, classroom teachers and students need to learn about using anti-virus scanners and virus deleters (shareware or registered). Excellent sources are found at the following World Wide Web sites: http:/www.brs.ibm. com/ibmav.html, http:/www.comm and.com, http:/www. mcafee. com, http:/www. symantec.com/avcenter.

It is helpful to preview several highly recommended programs (available at most of the web sites listed above) before investing in one. Fortunately, there is expert advice available, as well as software programs that will prepare the computers in your homes, schools, and offices for safer web surfing and Telnet commuting. Ways to reduce the chances of contracting computer virus infections include developing responsible use procedures, such as: 1) boot (start) your computers with licensed floppy disks (not pirated copies), 2) use only software that you've scanned with anti-viral software after downloading from legitimate bulletin boards, 3) scan all student disks before transferring data, 4) scan E-mail, 5) write-protect (close one of the top tabs) on virus free diskettes that have information that you don't need to update, and 6) use current anti-virus software.

Although computers with virus infections are not amusing, sometimes jokes are remembered longer than stern warnings to scan home, office and school computers:

- Politically Correct Virus: Never calls itself a "virus," but instead refers to itself as an "electric micro-organism."
- Government Economist Virus: Nothing works, but all your diagnostic software says everything is fine.
- Federal Bureaucrat Virus: Divides your hard disk into hundreds of little units, each of which does practically nothing, but all of which claim to be the most important part of your computer.
- Gallup Virus: Sixty percent of the PCs infected will lose 38 percent of their data 14 percent of the time (plus or minus a 3.5 percent margin of error).
- Congressional Virus: The computer locks up, screen splits erratically with a message appearing on each half blaming the other side for the problem (Missouri.edu, 1996).

These are just a few memorable reminders to start with . . . and, in no time, you'll be updating your classroom and/or office with new reminders rather than exotically named infections. With thoughtful planning school administrators, classroom teachers, students and their parents will benefit from computer technology without hearing . . . "The computer virus ate my homework!"

References

Symantec AntiVirus Research Center. (1996). What's New In the NAV Definition Update. Available from http:/www. symantec.com/avcenter.

Missouri.edu. (1996). Computer Users Beware of These New Computer Viruses: Virus Alert. Available from http://www.missouri.edu/~c639692/images/bgrpaper.gif.

Providing Medical Education Resources Over the World Wide Web

Joan Freedman M.S
Johns Hopkins School of Medicine, Baltimore, MD, USA
freedman@welchgate.welch.jhu.edu

Harold Lehmann, M.D., Ph.D.
Johns Hopkins School of Medicine, Baltimore, MD, USA
freedman@welchgate.welch.jhu.edu

Background

Medical education differs from most higher-education programs in that during the first two years of training students take a set-sequence of basic science lectures and labs. The third and fourth years are spent rotating through an elective sequence of clinical and inpatient services throughout the city. Due to the unique nature of learning medicine through patient care, it is difficult to insure all students exposure to patients with all diseases. As it is important to insure that students have timely access to educational materials to supplement the clinical experience, the World Wide Web was the logical choice for distributing educational materials.

LectureLinks is a distance learning framework comprising Web links that assist students at Hopkins in accessing educationally relevant resources that exist both within and outside of the institution. For each course lecture notes, old exams, Web-based CAI (computer-assisted instruction) materials were gathered and uploaded to the file server. In addition to resources created at Hopkins, students have identified external links to sites that contain materials relevant to their studies. By collecting these resources and organizing them in searchable, lecture-by-lecture format, students are exposed to a wider range of educational materials than they would normally have time to seek out.

This project was created as a response to students plea for relevant, timely computer-based programs to enhance and assist their learning and to provide options other than those assigned by the faculty. Since medical students have great demands put on them, they often do not have time to ìsurfî the Web or to seek out relevant and appropriate software resources on their own. Moreover, faculty members have used LectureLinks to see what other faculty members are teaching. Over the twelve months that LectureLinks has been running, it has become the de facto distance learning framework for the medical school curriculum.

Design

Begun as a series of manually managed HTML files, the LectureLinks framework has been redesigned to make it generalizable and to be usable in any higher education discipline. A set of databases and server scripts have been created to facilitate and to automate the updating process of LectureLinks (or any set of Web pages with many changing links). These scripts automatically update the database and recreate the HTML pages from the database.

The kernel of this design is the recognition that LectureLinks represents a database that matches resources to educational activities (see Figure 1). The database reflects this distinction by representing information about activities separately from resources. Activities can be nested, or be part of other activities (e.g. a lab is part of a course which is part of a year.). Resources can be local, as in lecture notes, or remote, as in Web sites at other institutions.

Information about the curriculum is entered into the LectureLinks Educational Activity table and updated as courses, lectures, and schedules change. This table contains the following information: activity type (lecture, seminar, lab, etc.), title, title as it will appear on Web page, faculty member and contact information, dates offered, keywords.

Relevant Web sites are identified either by the students, faculty, or the Office Of Medical Informatics Education and are submitted to the Update database via Web-based forms. Updating scripts are run periodically to add new URLs to the LectureLinks Resource table which contains the following information: resource type (e.g., tutorial, lecture notes, etc..), URL, title, finder's identity (who suggested the link), verification information, current status (include or not), content-based keywords.

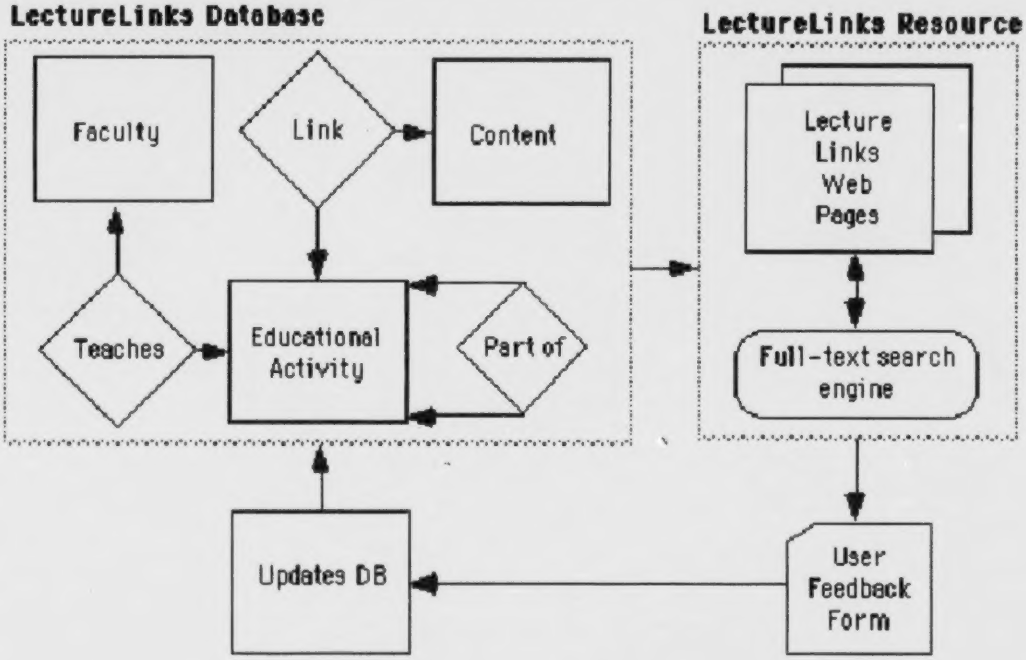


Figure 1: Entity relationship and data flow diagram for the LectureLinks project. The LectureLinks database is comprised of four tables. The Updates database is comprised of one table.

A link table joins the resource and activity tables by matching a URL to an educational activity. An activity may be linked to many resources, and a single resource can be linked to many activities. This table contains the following information: activity ID, resource ID, rationale (why activity and URL have been matched).

As often as is needed, scripts are run to recreate the LectureLinks pages to include new links and to remove unnecessary links. These scripts create updated and properly marked-up files ready for uploading to the server.

Assessment

An implicit part of the design of LectureLinks includes Web-based resource evaluation forms. Users can automatically submit feedback on which links are useful, which are not, and which new links should be added. Saving, sorting, and compiling this feedback will give a good picture as to the usefulness of the links. Additional data will be gathered through a student survey at the end of each academic year.

An additional assessment of use will be made by tracking numbers and types of local logons. By keeping track of the IP numbers used to visit LectureLinks, we can see where and when the students are logging on. Sorting logons by PPP/home access, Information Resource Center, or other campus/hospital

BEST COPY AVAILABLE

sites enable us to evaluate the hypothesis that access to curricular materials should be free of time and location restraints.

LectureLinks can be accessed at
http://infonet.welch.jhu.edu/~omie/LectureLinks/LectureLinksContents.html

Global Distance Learning and the Transformation of Knowledge Construction

Molly Freeman
ABD
The Union Institute
Lecturer, Dept. of Human Development
California State University
Hayward, California, USA
mollyfreeman@telis.org

Margaret Wheatley says, "I would like to work in an organization where I am so well trained that I can go (into the field) and know what to do under any circumstance...."[Wheatley,95] This summer I participated in training 1,000 K12 educators in Santa Clara County, the center of California's Silicon Valley. We trainers were trained so well, over periods varying from one to three years, that we did know what to do. We knew the fundamentals of appropriate use for learning with technology: 1)personal networks, 2)curriculum integration, 3)Internet skills. We knew that the Internet in the curriculum transforms the teaching process to facilitating, guiding, observing, listening; that it is a catalyst for self-motivation and creativity, and that it invites collaboration, cooperation and co-learning. We knew that patience, humor, practice and reiteration were keys to the mastery of new skills.

The Web, or Internet, became the occasion and the medium for new experiences and images of learning and creating knowledge. In the five day training we enabled participants to: 1)forge personal realtionships around their interests in curriculum and teaching stragegies, 2)acquire Internet navigation skills including Web, Gopher and Telnet searches, Email management and basic HTML and 3)conceptualize learner-centered curriculum projects. An additional three days led by professional facilitators focused entirely on collaborative curriculum projects.

Common denominators among the participants were: 1-interest in mastering Internet navigation skills, 2-anxiety about the project development process, and 3-uncertainty regarding the technical assistance, connectivity and access to computers in their respective schools for the ensuing year. Frustration, anger, some tears and despair were closely juxtaposed with determination, persistence, timidity and elation as familiarity with the technology increased from Day 1 to Day 5. People who had no prior experience with the Internet on Day 1 were planning Internet integrated projects by Day 5 and elaborate, collaborative learner-centered projects by Day 8.

Complexity theory has a way of explaining how this happened.

The Web captures the imagination, it compels us to stay tuned, clicking and searching, replying and expounding, mixing and relating. The Web is the ongoing report of interdependent information processing systems, humans and computer code, combinations of zeroes and ones with meanings, values and perceptions. A part of this attraction is the novelty of the Web and the ongoing novelty it makes possible through HTML linking and site building, through color, sound and image with text.

Participants in the Summer Institute were told that they were helping to determine how education would be transformed by the new technology, and within two days they came to experience the diversity of the Web itself. They experienced themselves as information providers and users, as senders and recipients of email. They were hooked because they were intimately involved. Hands-on activities included scavenger hunts where the labs became hushed, with silences punctuated by "I got it." Email retrieval activities were replete with "I got mail!" "Did you get mine?"

We set forth almost 1,000 teachers prepared to utilize the Internet with classroom curricula across five counties in northern California because people experienced themselves as interdependent and vital to change. People experienced themselves as active learners enticed by the opportunity to be on "the cutting edge" as contributors, so work was transformed to voluntary participation with responsibility.[Lee, 59] Self-motivation was activated by the energy released from learning and acquiring new competency.

The Summer Institute96 succeeded in preparing teachers to envision and plan Internet integrated curricula because we worked off the energy released by the Web. We had the extraordinary opportunity to introduce people to this new way of accessing information and one another, to a new way of being in the world. We gave them discrete activities that provided a sense of accomplishment and an awareness of themselves as learners, analogous to what Paolo Freire means when he speaks of enabling people "to name their world."[Freire,70] We enabled K12 educators to "navigate the new world of information."

The new competency releases enormous creative energy for actualizing the curriculum projects, which may indeed create considerable turbulence within classrooms, schools and districts. From the perspective of complexity theory this is as it should be. Indeed, there is a concept taken from chemistry called the "molecular clock," which refers to a condition of turbulence when molecules in a substance change, all at once.[Prigogine & Stengers, 83]. I contend that one way the Web is altering our experience of learning and of education is through our adaptation to the frequent transformation of images and text that we witness on the computer screen. It is as if we were witnessing and participating in multiple "molecular clocks" each time we are online.

In another learning environment, an amphitheater lecture hall on a California State University campus, where the students sat at a 60 degree angle to me, I witnessed the dynamics of transformation similar to those in the Internet training, albeit minus the Web. The course was Aging as Lifelong Learning, where 40 students participated in face-to-face lecture and conversation and 130 participated via one-way audiotapes of the inclass proceedings. Students could email, leave voice messages, fax or use the US postal service. I responded by voice, fax, email and US post.

I contend the transformation of learners was similar because the message of learners as knowledge creators was the same. The common thread between the K12 Internet training and the amphitheater, audiotape course was the intention of presentation and of the readings; Complexity: The Emerging Science at the Edge of Chaos[Waldrop 92] and We Are All Self-Employed[Hakim, 96]. The intention being the pursuit of authentic dialogue accompanied by explicit admonitions to students to understand their own internal conversations as the self-referencing of interdependent, adaptive persons participating in the creation of knowledge about the world and about their own futures.

Similarities between these two distinct teaching experiences resonate with lessons from complexity theory to listen to and observe the learning, the self-referencing, the knowledge-creating of students, and, by so doing to honor them. Transformed by the honoring intention, students experience themselves as active, interdependent participants in the world...with a keen sense of responsibility for what they do with the thoughts they are building and refashioning. They learn to value and to attend to their own thinking processes.

This image of learning and teaching is not new. It is continuous with the history of education in Western Civilization, from Plato to Rousseau, to John Dewey, Martin Buber and Paolo Freire...along the way are Maria Montessori, Jean Piaget and Sylvia Ashton-Warner and Reuven Feuerstein, to name only a few. Learning proceeds through experience and doing and through knowing thyself, in relation to the world.

In short, the World Wide Web, as well as many practitioners of global distance learning, [Bensusan,95], [Perrin,95], [Lane,94], [Hiltz,94], actualize the dialogical education and mediated learning experience that philosophers and educators have been describing for centuries because the Internet as Web and as real time interactive conferencing software, are quintessential information processors. In the language of complexity theory this is what complex adaptive systems do....they process information with their environments in terms of their own characteristics. So, interdependence is continuous self-referencing of different sets and kinds of information. The magic of the Web, and the interactive capacities of the Internet, what make them so compelling, is their capacity to engage our self-referencing. It hooks us precisely because it is so familiar, so much like what we do as we coexist, coevolve with our environments and with one another.

This similarity in activity or play is one explanation of the prolific work of the MIT Media Lab, and an argument for high technology industry to adopt schooling as a giant test bed, contributing to education as a form of corporate R & D. This, in turn, brings us back to the construction of knowledge issue, to the creation of knowledge as we learn, on the edge of chaos, where we remain poised to unlearn what we may know best in order to learn new things that will make the system better.[Waldrop,92] John Holland speaks of this feature of complex adaptive systems and of the recombination of building blocks (eg. *public-private partnerships between schools and industry*) as systems self-organize with changing environments. [Holland, 94]

The language of complexity theory can sustain the employment of knowledge workers as it helps educators come to terms with our transformation from being sources of knowledge to acting as facilitator, guide, docent, colearner and researcher. As the boundaries between teaching and learning dissolve educators become more like participant observers, coming to terms with the interdependence of our own theories and praxis with where, when and what we observe. At the same time new praxis emerges around teaching as honoring when we learn the "world is a matter of patterns that change, that partly repeat, but never quite repeat, that are always new and different."[Waldrop,92] Or, in awe, we approach each class, each colearner as one approaches sacred text: anticipating new insights and understandings, to be read each time as if for the first time.

References

[Bensusan,95] Bensusan, Guy (1995). Interactive Instructional Television: The Bensusan Method. ED, Education at a Distance, November 9 (11), 5-14.

[Freire 70] Freire, Paolo (1970). Pedagogy of the Oppressed. New York: The Seabury Press.

[Hakim 96] Hakim, Cliff, (1995). We Are All Self-Employed. San Francisco: Berrett-Koehler Publishers.

[Holland 95] Holland, John H., (1994). Hidden Order, How Adaptation Builds Complexity. New York: Addison-Wesley Publishing Company.

[Hiltz 94] Hiltz, Roxanne Starr (1994). The Virtual Classroom: Learning Without Limits Via Computer Networks. NJ: Ablex Publishing Corporation.

[Lane,94] Lane, Carla (1994). The Role of Technology in the Systemic Reform of Education and Training. ED Journal, June 8 (6), 1-22.

[Lee 59] Lee, Dorothy, (1959). Freedom and Culture. Englewood Cliffs, NJ: Prentice-Hall Inc.

[Perrin,95] Perrin, Elizabeth (1995). The Continual Battle in Distance Learning. ED, Education at a Distance, March 9 (3), 6.

[Waldrop 92] Waldrop, M. Mitchell, (1992). Complexity and the Emerging Science at the Edge of Chaos. New York: Simon & Schuster.

[Wheatley 95] Wheatley, Margaret J. (1995). Keynote Speaker, Healthy Communities Summit, Healthcare Forum, 1995, San Diego, California.

Acknowledgements

I am grateful for the opportunity to have worked with the SCCOE-SJEN Summer Institute96; Judy Powers, Institute Co-Director and Manager, Education Technology Curriculum; Diane Berthoin-Hernandez, Summer Institute Training Coordinator. Santa Clara County Office of Education, 1290 Ridder Park Drive, MC #237, San Jose, California 95131-2398. For inquiries: Judy Powers, jpowers@telis.org, voice: 408-453-6523, fax: 408-453-6875.

I thoroughly enjoy the opportunity to learn from fauclty and students in the Dept. of Human Development, California State University, Hayward, Hayward, California

Homeland: An Integrated Set of Tools to Support Regional Telecommunities

Nicola FREGA
GIUDA Lab.
University of Calabria, ITALY
nfrega@giuda.deis.unical.it

Antonio VOLPENTESTA
GIUDA Lab.
University of Calabria, ITALY
volpente@giuda.deis.unical.it

Tommaso FASANO
GIUDA Lab.
University of Calabria, ITALY
tfasano@barabba.deis.unical.it

Antonio GRECO
GIUDA Lab.
University of Calabria, ITALY
agreco@giuda.deis.unical.it

Introduction

Four years ago, at Giuda Lab., we've decided to develop a BBS for educational purpose. The system was intended to support educational activities inside the campus of University of Calabria. Actually, during last years, we've been observing a natural growth of user community spread all over Cosenza hinterland, not only interested in educational issues.

With the increasing number of users (at the time being more than thousand) new needs have arisen up. This has lead to extend BBS local boundaries towards a national and international scenario. *Homeland* has been designed as technological solution to realize such an extension.

Currently *Homeland* is used as a virtual square where people interested in places, traditions and customs of Calabria may meet and exchange information. In addition to that it has allowed many calabrians, especially those who live in foreign countries, to get in touch and keep in contact with their relatives and friends.

Architectural Issues

Homeland is a system of tools that integrates Internet and conventional BBS technologies. In particular it essentially consists of some Internet modules, BBS modules and interfaces between them.

On the Internet side we've built-up:

- *Mailing List* : to broadcast e-mail messages among Internet Homeland subscribers;
- *News server* : messages sent to Homeland are made available and classified as local newsgroups;
- *Web Board* : to let users browse all messages and bulletins from WWW;
- *MBBS* (Multimedia BBS) & *Postcard Service* :after retrieving an image from a local Pictorial Database (PDB) and adding it text, users can post a multimedia message to a Public Web Board or to a private mailbox; images in PDB deal with Calabria and the retrieval software is based on

many different techniques.

On Giuda BBS side we have:

- *Netmail* : personal mailboxes for BBS users;
- *File Areas* : archives of binary files such as images, documents and programs;
- *Conferences* : public thematic message areas.

The BBS-Internet communication has been achieved by a certain number of gateways (shortly GW):

- *Mailing list GW* : to enable message exchanges between the Internet mailing list and dedicated BBS conferences;
- *News GW* : to interface BBS conferences with the newsgroups;
- *File GW* : to allow BBS users to send images to Homeland PDB and retrieve them via BBS.

In what follows we outline the *Homeland* architecture with respect to the Internet side ; afterwards we take a closer look at it by focusing our attention on the Web module.

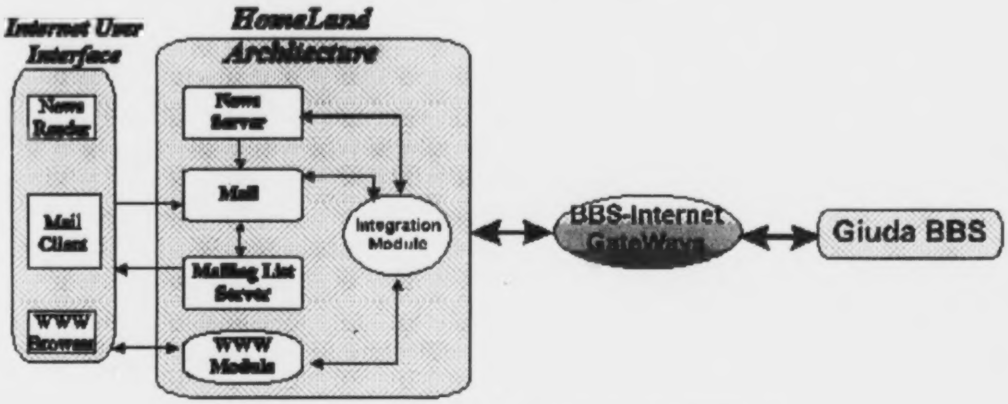

Internet User Interface
HomeLand Architecture
News Reader
Mail Client
WWW Browser
News Server
Mail
Mailing List Server
WWW Module
Integration Module
BBS-Internet GateWays
Giuda BBS

Figure 1 : HomeLand architecture : Internet side

The *integration module* is devoted to collect heterogeneous kinds of incoming message (news, mail, html), to transfer them in suitable format to each inner module and to forward them to the BBS-Internet gateways. In such a way user can communicate each other even if the use different net clients.

The logical structure of WWW module is showed in the following figures.

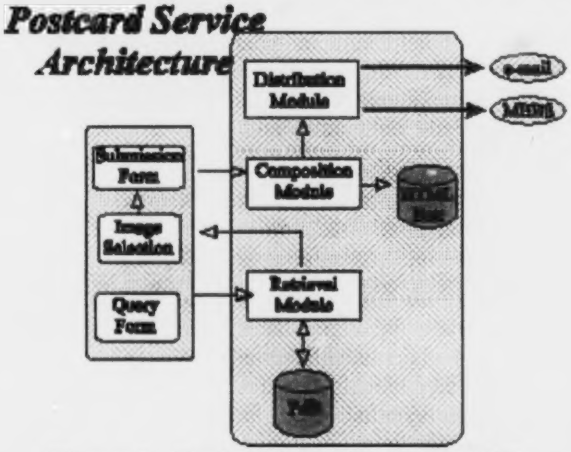

Postcard Service Architecture
Distribution Module
e-mail
MBBS
Submission Form
Image Selection
Query Form
Composition Module
Retrieval Module
PdB

Figure 2 : Postcard Service and MBBS

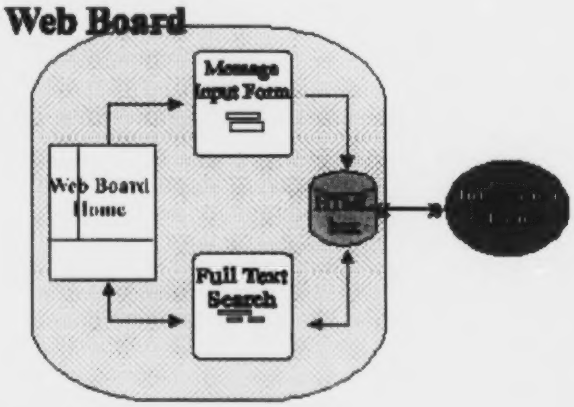

Web Board
Message Input Form
Web Board Home
Full Text Search

Figure 3 : the Web Board

BEST COPY AVAILABLE

References

HomeLand can be reached at *http://barabba.deis.unical.it/homeland/*. An extended version of this paper can be found at *http://barabba.deis.unical.it/Documents/WebNet96/homeland.html*. All the following referenced papers are at *http://barabba.deis.unical.it/Documents*

[Frega and Volpentesta 94] Frega N., Volpentesta A. (1994). A Multimedia Bulletin Board in the WWW environment, 2nd International WWW Conference Mosaic and the Web, 1994, Chicago IL.

[Greco 95] Greco A.(1995), GIUDA BBS Tre anni di esperienze in un Campus universitario e dintorni, Reti Civiche ed Autostrade Digitali, 1995 Rende, ITALY

[Frega 95] Frega N. (1995), Agenti, Robots e dimostratori per reti civiche, Reti Civiche ed Autostrade Digitali, 1995 Rende, ITALY.

[Vopentesta 95] Volpentesta A. (1995), Telecomunità tra il senso della realtà e il senso della possibilità Reti Civiche ed Autostrade Digitali, 1995 Rende, ITALY

A Web Based Course on Dreams: Teaching Personally Sensitive Material Over the Internet

Jayne Gackenbach, Ph.D.
Computer and Outreach Services
Grant MacEwan Community College, Canada
jgackenb@gpu.srv.ualberta.ca

Unlocking the Secret of Your Dreams is a noncredit introductory course that is delivered on the World Wide Web by Grant MacEwan Community College. The course consists of five lessons which teach the student to interpret and work with their own dreams, both alone and in electronic group sessions. Each lesson is accompanied by an assignment where students put what they have learned into practice, and answer a questionnaire about their dream experiences. Students receive feedback on the completed assignment the following week.

Here is a brief outline of the lessons and assignments. A longer version is available on the registration page for the course (http://www.outreach.org/dreams):

Lesson #1 - Introduction to Dreaming

Summary: The latest psychological, clinical and cross cultural scholarly work into these remarkable experiences of the night are offered so that the student can develop a working understanding of the process of dreaming. How to recall dreams and techniques for dream diarying are covered.

Assignment: This is a basic demographic questionnaire including questions concerning their sleeping and dreaming habits and preferences.

Lesson #2 - What Do Most People Dream About?

Summary: The typical content of peoples dreams is presented followed by techniques for ascertaining the content of ones own dreams relative to norms. It is stressed that the dreamer and not the course instructor, friends, psychotherapists, scientists or anyone else is the owner of their dream and thus the final authority on his/her dream.

Assignment: The dreamer is asked to answer various questions about their dream experiences.

Lesson #3 - Dreamwork Tools & Techniques

Summary: In this lesson the role of metaphor in interpreting dreams is stressed. Basic dreamwork tools are covered which do not rely on an expert.

Assignment: The student is asked to work one of their own dreams in the privacy of their home first in a dream metaphor exercise. Then they are asked to use an individual dreamwork technique to further interpret their dream.

Lesson #4 - Dream Connectedness

Summary: Our culture often thinks of the dream as a highly personal experience. Although it is this, there are universal themes which resonate for peoples from around the world. These themes have been discussed by various dream theorists and are covered in this lesson.

Assignment: Students are given the option to join dream groups in bulletin board format. A step by step procedure is outlined based on the lesson material encouraging these small groups to engage in group dream interpretation.

Lesson #5 - Selected Topics

Summary: These topics include some discussion of intensified forms of dreaming such as nightmares; characteristics of dreams in specific populations such as those addicted to a substance or in children.

Assignment: Students are offered the opportunity to receive a list of e-mail addresses of others in the class who are interested in pursuing group dreamwork on the Internet. They are asked to evaluate the various components of the course.

Also available on the dream course home page is a information related to the course. This includes: a student bulletin board where students post messages about the class or about their particular interest in dreams; a list of recommended books and links to dream related websites (an annotated list of links follows a qualifier about linking to other dream sites on the Internet. This gives the student some idea of what else is available and addresses the potential problems with accessing sites which offer dream interpretation.); a course bibliography; biographical information about the instructor, including her photograph and a link to her full curriculum vitae; and e-mail dialogue boxes to contact the instructor and/or the course web master.

This course is about 170 pages of written material with about 35% of that as assignments. Because the course is offered on a web site, after an introduction to the unit the student is presented with a table of contents for the topics included in that unit. Thus they can read the unit linearly or skip picking which topics are most interesting or relevant for them. To enhance a sense of personal authority and psychological safety, throughout the course material it is stressed that the dreamer is the final authority on their dreams.

In addition to the demand to write in relatively fully contained idea units, the level of the writing had to remain rather elementary because of the potential target market. Also much of the writing was done in the first person in order to enhance the felt experience of the student of being in a dialogue with the instructor. Thus personal examples from the authors own life and work were used in an effort to personalize the instruction.

The assignments were designed as primarily closed ended questionnaires with ample opportunity for commentary and dialogue. Because of this format the shape of the students information is quickly fed back to them. Additionally, this form of response appeals to students in terms of increasing the likelihood that they will initially fill out the questionnaire and thus continue with the course. As they become more comfortable with the material they tend to prefer more open ended items. An important contribution of a web site based assignment is the opportunity to take advantage of hyperlinking to increase the information possibilities and structural potentials for learning. For instance, when filling out questionnaires people are often curious about how others may have answered a particular item. In some assignments the student can link to such normative data. Although from this may bias the students response it helps, especially with sensitive material, to encourage the student to realize that their experiences are not as unusual as they often fear when working with dreams.

For the fourth assignment the students are given an opportunity to participate in an on-line dream group. The model of these groups is leaderless and without theoretical assumption. After acknowledging that they have read a series of 'groundrules' which not only explain the procedure of the groups but also the ethical aspects of working in a dream group in any setting, the student is assigned to a dream group. The form of these on-line dream groups is using a bulletin board to which they can post their responses to the dreams of their classmates. These dream group postings are monitored by the instructor to insure the psychological appropriateness of the responses within the loose structure. Suggestions for appropriate posting on a bulletin board as well as the basic outline of the group procedure are available to the student each time they access their dream group bulletin boards. The students can participate anonymously, simply read postings without posting themselves or give details about themselves in the text of their postings. This allows for a wide variety of self-disclose as is comfortable for the various participants.

Acknowledgments

Although I am listed as the author and expert of this course on dreams, bringing it to the Internet community was in reality a team effort. I would first like to thank Grant MacEwan Community College who provided the team whose vision and enthusiasm for the project made it happen. Our "Dream Team" included Russ Powell, Fred Sawka, James Morrissey, Kevin Collard and Greg Farris.

Implementing Engineering Education on the WWW: 3 Case Studies

Gale, Carolyn M. Center for Innovation in Engineering Education, Vanderbilt University, USA. galecm@vuse.vanderbilt.edu

Bourne, John R. Director, Center for Innovation in Engineering Education, Vanderbilt University, USA. john.bourne@vanderbilt.edu

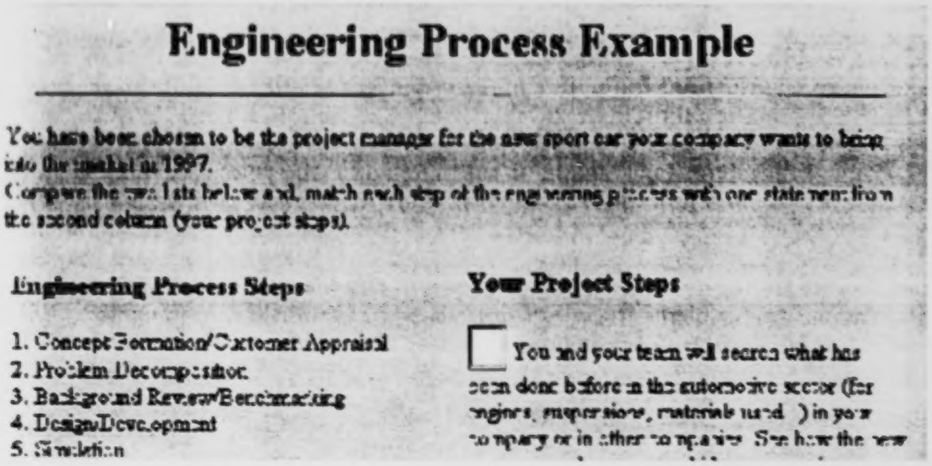
Dawant, Martine M. Research Instructor, Center for Innovation in Engineering Education, Vanderbilt University, USA. datares@ctrvax.vanderbilt.edu

Quinn, Doris. Director, Quality Education and Measurement, The Vanderbilt Clinic, Vanderbilt University, USA. quinndc@ctrvax.vanderbilt.edu

Faculty and students of the Vanderbilt University Center for Innovation in Engineering Education have developed several complete on-line courses primarily for use by students of the Vanderbilt School of Engineering. Three of these courses will be discussed: Introduction to Engineering, Informatics, and Total Quality Management.

Introduction to Engineering (ES 130)

ES 130 is an introduction to engineering computer tools for freshman engineering students. This course was designed to introduce basic computer knowledge, the Internet and information resources, modeling and simulation methods, mathematical computer tools, and team work. Students were able to access all lectures and interactive laboratory sessions on-line. Many of the lectures included software tutorials and additional interactive examples.



Engineering Process Example

You have been chosen to be the project manager for the new sport car your company wants to bring into the market in 1997.
Compare the two lists below and, match each step of the engineering process with one statement from the second column (your project steps).

Engineering Process Steps

1. Concept Formation/Customer Appraisal
2. Problem Decomposition
3. Background Review/Benchmarking
4. Design/Development
5. Simulation

Your Project Steps

☐ You and your team will search what has been done before in the automotive sector (for engines, suspensions, materials used) in your company or in other companies. See how the new

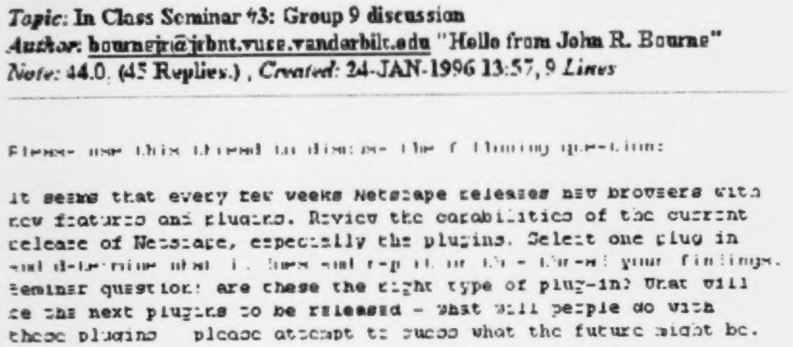
Figure 1. ES 130 included interactive examples in the lectures as well.

ES 130 can be found at http://ciee.vuse.vanderbilt.edu/es130/home130.html

Informatics (EE/MT 274)

BEST COPY AVAILABLE

Informatics is a course designed for Electrical Engineering/Computer Science undergraduate and Management of Technology graduate students. The outcome of this course was to gain an understanding of information technology and its resources. While many course materials could be found on-line, the primary focus was a Web-based conferencing system that required students to post and discuss all class assignments and team projects.



Topic: In Class Seminar #3: Group 9 discussion
Author: bournejr@jrbnt.vuse.vanderbilt.edu "Hello from John R. Bourne"
Note: 44.0 (45 Replies.) , *Created:* 24-JAN-1996 13:57, 9 *Lines*

Please use this thread to discuss the following question:

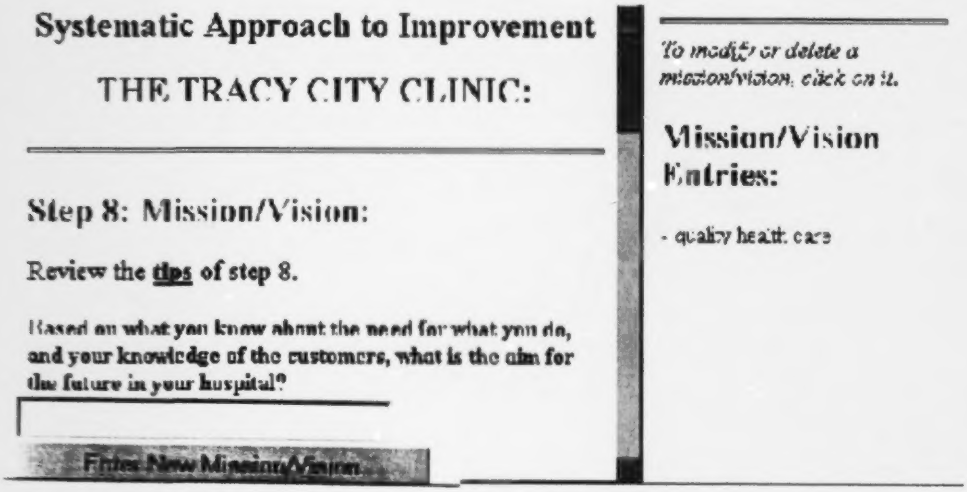
It seems that every few weeks Netscape releases new browsers with new features and plugins. Review the capabilities of the current release of Netscape, especially the plugins. Select one plug in and determine what it does and report on it - discuss your findings. Seminar question: are these the right type of plug-in? What will be the next plugins to be released - what will people do with these plugins please attempt to guess what the future might be.

Figure 2. A seminar topic posted by the instructor for EE/MT 274.

EE/MT 274 can be found at http://jrbnt.vuse.vanderbilt.edu/

Total Quality Management (MT 322)

Total Quality Management (MT322) is a graduate level course that teaches the fundamentals of total quality measurement. The curriculum includes case studies in manufacturing and in the health care fields. A quantitative approach is emphasized using various quantitative tools for measure quality metrics. Students use the on-line materials to submit homework and to discuss questions posed by the instructor.



Systematic Approach to Improvement

THE TRACY CITY CLINIC:

Step 8: Mission/Vision:

Review the tips of step 8.

Based on what you know about the need for what you do, and your knowledge of the customers, what is the aim for the future in your hospital?

To modify or delete a mission/vision, click on it.

Mission/Vision Entries:

- quality health care

Figure 3. An on-line assignment from MT 322.

MT 322 can be found at http://ciee.vuse.vanderbilt.edu/quality/home322.htm

BEST COPY AVAILABLE

Reorganised by the Web: Bringing the Missing Link into Focus

Paula Garrett, University of Sydney Library, Australia Email: p.garrett@library.usyd.edu.au

Steve Ryan, University of Sydney Library, Australia Email: s.ryan@library.usyd.edu.au

The Web environment is a "natural" for the delivery of many types of information based client services. In libraries the Web offers a central platform from which a wide and increasing range of information resources can be organised as well as accessed. Moreover, the Web provides the most suitable vehicle for keeping up with the continual change inherent in the new information and technology marketplace and the digital library. Web sites will probably always be "under construction." Adaptability and visibility are increasingly becoming our corporate and academic survival tools.

Information based organisations such as libraries can actually use the process of designing a client-focussed Web as a restructuring tool if staff resources are properly allocated or reallocated along the way. The Web design process forces an organisation to take a hard look at values, priorities and processes. Mimicking library departmental structure--Acquisitions, Cataloguing, Circulation, Reference--on the Web will most often be meaningless and useless to clients. Web design needs to center around client processes not staff procedures. In this way the Web can begin to graphically reflect the significant changes needed in an organisationís structure, particularly in its staffing, to better support its clientele.

To date there are library Web sites, such as the University of Sydney Library [HREF1], exemplifying the framework of a client based interface, but often the substance or content is thin. Typically, Web designers and authors will add those activities on top of their existing responsibilities. Then the Web begins its evolution as a useful, valuable organisational presence. Expectations grow, workloads increase, and management issues inevitably surface. In pyramid shaped institutions Web development often stays isolated with a few staff members. Most likely this isolation will bring about its demise and thereby the failure of these information organisations to adapt to the demands of their changing service environment.

In creating a client focussed Web design, new collaborations among staff and between staff and clientele emerge if the structure is supportive. Collaborations encourage greater information sharing and result in what Paepcke refers to as ìinformation compounds,î new information ìconstructed from the accumulated pieces of previously retrieved and newly acquired information.î [Paepcke 1996 [HREF2]] Staff and clients ìneed to find, analyze, and understand information to use it in multiple contexts, and to manipulate it in collaboration with colleagues of different backgrounds and focus of interest.î [Paepcke 1996 [HREF3]]

Organisations will probably find that the intersecting circles of staff which form in designing a Web site will not fit into their pyramid shaped structures. Web shaped structures of management have been discussed by authors like Sally Helgesen in *The Web of Inclusion.* Although Helgesen does not use the World Wide Web model, her concepts closely resemble those upon which the World Wide Web is based -- collegial participation, open communication and accessibility. ìSince web structures are circular rather than pyramidal, those who emerge in them as leaders tend to be people who feel comfortable being in the center of things rather than at the top, who prefer building consensus to issuing orders, and who place a low value on the kind of symbolic perks and marks of distinction that define success in the hierarchy. This preference on the part of web-style leaders infuses their organizations with a collegial atmosphere, which in turn enables people to focus upon *what* needs to be done rather than *who* has the authority to do it.î [Helgesen 1995]

Letís look at the contrasting qualities and attributes between the traditional pyramid and the evolving web as models of organisational structure.

Pyramid Web

Rigid Hierarchies Flexible Teams

Isolation of Power Shared Responsibility

Fragmentation Integration

Control Open Access

Rigid Hierarchies vs. Flexible Teams

Top-down organisations are usually not very agile when it comes to responding to and more importantly, managing change. And change is now a constant in the information marketplace. Flexible teams which interact to develop a libraryís Web will be better situated to perceive change in client groups and the industry and can thereby incorporate it as necessary. Fluid boundaries support interdisciplinary projects, the structure flattens and all participate in the creation of an enriched environment of discovery and disclosure.

Isolation of Power vs. Shared Responsibility

The concentration of power in a pyramid is at the narrow top. Quite often in these structures decisions have to make their way from the bottom up through each hierarchical layer until they sometimes find their way to the top by which time of course the plot may well be lost. In using the working group or team model which succeeds in the process of Web development, responsibility gets shared across the structure. Senior management must lead the organisation with a client based focus, ensure that they hire capable staff, train and retrain when required, and then let the staff get on with it.

Fragmentation vs. Integration

Traditional departmental structures have segregated processes in the name of procedure, and often these procedures are based on extinct systems such as a card catalogue in a library. More importantly, departmental structure has in many cases fostered destructive competition, inequities and ultimately, the clientele has been lost from sight. As systems or IT units have been introduced in libraries, they have typically been set up as yet another department. Technology is no longer a trend or phenomenon; it needs to be seen as just one of a number of components which are seamlessly integrated into the operations of information based services and digital libraries. Whatís needed is iinformation systems research, instead of computer technology research. Digital libraries need to be tested with large collections and users since the value of the technology cannot be evaluated in isolation." [Schatz 1995 [HREF4]]

Control vs. Open Access

Control from the pyramid ìtowerî may have worked or at least appeared to work when the main business of information based services such as libraries was handling physical objects. In those days information was a finite set of items. In the digital library we have moved to managing bits not atoms [Negroponte 1995]. Management is now responsible for making accessible to its clientele an expanding, widely distributed and very complex network of information objects including databases, software, interfaces, metadata and people. Not only do the clients need an open access structure, but also the staff will have to have such an environment in order to move the organisation forward.

Libraries have been utilising technology for decades now, but the developments in the distribution of information via the Internet and more specifically the World Wide Web have profound implications including major shifts in the responsibilities of libraries. Technology will have to be truly integrated and managed by the content providers--collaborative teams of staff and clientele. Divorcing the process of Web development from the content providers of information based services will expose an organisationís lack of relevancy to clients and staff, its user community.

A web based structure will need to stay open and dynamic. Like Web sites, the organisation will continually be ìunder construction.î Using the process of Web design as both a tool and a mirror, organisations can naturally progress -- working through trial and error with integrity and cohesiveness between structure and process.

References

[Helgesen 1995] Helgesen, S. (1995). The web of inclusion. New York, Currency/Doubleday.

[Negroponte 1995] Negroponte, N. (1995). Being digital. Rydalmere, N.S.W. : Hodder & Stoughton.

Hypertext References

[HREF1] University of Sydney Library. [WWW document] http://www.library.usyd.edu.au/

[Paepcke 1996 [HREF2]] Paepcke, A. (May 1996). Digital libraries: Searching is not enough: What we learned on-site. D-Lib Magazine. [WWW document] http://www.dlib.org/dlib/may96/stanford/05paepcke.html

[Paepcke 1996 [HREF3]] Paepcke, A. (May 1996). Digital libraries: Searching is not enough: What we learned on-site. D-Lib Magazine. [WWW document] http://www.dlib.org/dlib/may96/stanford/05paepcke.html

[Schatz 1995 [HREF4]] Schatz, B. (August 1995). Report of the library perspective working group, 1995, Interoperability, scaling, and the digital libraries research agenda: A report on the May 18-19, 1995 IITA digital libraries workshop. [WWW document] http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/part3.html

GENDERBENDER: Gender and the Human Computer User Interface

Gregory Patrick Garvey
Department of Design Art
Concordia University
Montreal, Quebec, CANADA
ggarvey@vax2.concordia.ca

When a potential user first approaches GENDERBENDER the attract mode displays the following:

"Are you really a man or a woman or a little bit of both?"

"Etes-vous un homme ou une femme ou un peu les deux?"

"Now you can be sure (or can you?)!"

"Maintenant, vous en tes certain!(l'tes-vous?)"

Although on the Internet no one may know you are a dog GENDERBENDER performs a much needed public service by minimizing the cognitive dissonance of gender confusion and subterfuge found in chat groups, MUDS, and MOOS. Inspired by standard psychological tests for gender and personality profiles and Alan Turing's test for Artificial Intelligence. GENDERBENDER allows a user to self administer a gender test. Based on the user's responses the "Computer Psychologist" will display the message, "You are a man!" or "You are a woman!" or "You are androgynous!" The "two player" version allows two users to view the responses of one another. Each in turn can guess the gender of the other player and whomever the computer psychologist agrees with is the winner(?)!. CUseeme teleconferencing makes it possible for the users to compare a video simulacrum with the assessment of the Computer Psychologist.

At the start of both the single player and two-player version the Computer Psychologist will display the first of a series of questions randomly selected from a possible total of sixty. The Morph-o-meter displays KENBY an androgynous 'virtual' figurine. As a user answers each question with yes, no or don't know the Morph-o-meter gives instant feedback on whether masculine or feminine characteristics predominate in the user's personality by morphing towards a identifiably male or female figurine. The Tile-o-matic will reveal each user's video image tile by tile for each yes response. For each don't know both the Morph-o-meter and the Tile-o-matic do not change.

GENDERBENDER (Release 1.0) was exhibited as part of the summer installment of Image du Futur in Montreal (May-September). GENDERBENDER Release 2.0 will introduce the two player Internet version. GENDERBENDER Release 3.0 will contain the additional feature of the creation of an online avatar that reflects the gender profile that the user gives it. The Self-Test allows the user to construct a personal gender profile of twenty masculine, feminine or neutral traits. Once created it can act as a gendered knowbot that will visit chat groups, perform searches and then report back to it's master(?) on its discoveries, experiences, exploits and perhaps provide a little black book for actual meat and flesh encounters.

GENDERBENDER is loosely based on the 1974 Bem Sex Role Inventory (BSRI). The help feature of the software offers the disclaimer that GENDERBENDER is not an actual diagnostic test. Of course individuals needing help or assistance with personal problems should seek the assistance of qualified psychological or psychiatric professionals. The BSRI is a self-administered 60-item questionnaire, containing a Masculinity scale and a Femininity Scale with 20 neutral items as filler. For either the self-test or the two player game the software randomly selects questions from the 60 items. Ellen Lenney describes the BSRI "assesses masculinity and femininity in terms of the respondent's self-perceived possession of positive personality characteristics having sex-typed social desirability."[Lenney 1986] Several examples are listed as follows:

Acts as a leader Affectionate Friendly

Aggressive Cheerful Happy

Ambitious Childlike Helpful

Table 1: Sample items from the BEM Sex Role Inventory

The so-called "Turing Test" for Artificial Intelligence was first proposed by Alan Turing nearly 46 years ago in [Turing, 1950). The popular notion of this test is that a 'user' (sic) types questions which are responded to via screen display by either another person or software programmed to fool the user into thinking it is a person. Software which achieves this is said to pass the Turing Test for Artificial Intelligence. Turing actually sidestepped the question "Can machines think?" by proposing what he called the "imitation game" which anticipates by some 40 years the gender crossing behavior of habitues of Internet chat groups(documented by writers such as Kevin Kelly and others).

GENDERBENDER is not intended to be a literal implementation of Turing's "Imitation game". Rather it uses Turing's proposal of three way interaction as a way to structure the two player game and then restricts the questions to the 60 items from the BEM Sex Role Inventory. It is instructive to quote at length a portion of Turing's original prescription for "the imitation game".

"The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:"

"C: Will X please tell me the length of his or her hair? Now suppose X is actually A, the A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be: 'My hair is shingled, and the longest strands are about nine inches long.' In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten,. The ideal arrangement is to have a teleprinter communicating between the two rooms. Alternatively the question and answers can be repeated by an intermediary. The object of the game for the third play (B) is to help the interrogator. the best strategy for her is probably to give truthful answers. She can add such things as "I am the woman, don't listen to him!' to her answers, but it will avail nothing as the man can make similar remarks."

"We now ask the question, 'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does the when the game is played between a man and a woman? These questions replace our original, 'Can machines think?' "

Perhaps the real question is 'Can humans think--critically?' When personality traits become reduced to and locked in algorithmic descriptions those chosen traits almost inevitably reflect the biases and clichs of what is considered 'normal.' Moreover the BSRI of 1974 is a kind of time capsule giving insight into how mutable notions of gender really are. This lends credence to the politically correct view that "gender is a constructed cultural artifact". The potential for misuse by health care professionals of such psychological instruments are considerable. The obvious example is the pejorative label of androgynous i.e. does not conform to the binary either or opposition of male or female sex. Having a mixture of gender traits does not conform to this dichotomy and therefore must be considered abnormal requiring corrective therapy. Woe to those deviant in-betweens.

Alan Turing himself suffered from this law of the excluded middle having been persecuted and imprisoned under homophobic laws in Great Britain which may have led to his suicide in the fifties. Ironically he perhaps as much as any one individual contributed to the success of the Allies over Nazi Germany as he is often credited with cracking the code of the Enigma Machine used to encrypt messages of the Nazi U-Boat command. An apocryphal account suggests that at the time he wrote of the "imitation game" he was undergoing forced hormonal therapy and was sprouting breasts. Under the circumstances he was no doubt acutely aware of the advantages of narrow band communication.

References

[Lenney 1986] Lenney, Ellen, (1986) Sex Roles, 11, 582-588

[Turing 1950]. Turing, A. M. (1950). Computing Machinery and Intelligence. MIND A Quarterly Review of Psychology and Philosophy Vol. LIX, No. 236, 133-159.

GENDERBENDER Two Player, Single Player software. Copyright 1996 by Gregory Patrick Garvey

Health Care Information on the Web: Model and Menace

Laurie Gelb
Department of Biomathematics
University of Texas M.D. Anderson Cancer Center, USA
laurie@biomath.mdacc.tmc.edu

Transcending traditional product lines of foods, drugs, devices, and providers, today's medical organizations "market" intangibles such as risk-assessment/reduction, self-care and self-triage, plan/network enrollment, and, more generally, appropriate utilization of the professional resources they provide. Thus, health care offers its constituents specific benefits for purchasing and utilizing their offerings in certain ways.

Ironically, an increasingly customer-centered industry must now control customer expectations and manage demand as never before. Thus, the importance of communicating specific differences between appropriate and inappropriate care-seeking and lifestyle behaviors has never been greater. Behavioral changes are more likely to derive from personal experience than through abstract reasoning. The Web permits the integration of multimedia learning tools, updated information, a range of reputable sources and easy access to the histories of others. However, it often lacks a means through which to integrate an individual's characteristics and experience into Web-based health-related information-seeking.

For example, few sites offer any of the following: detailed, interactive algorithms for evaluating providers and plans, comprehensive provider directories, simplified explanations of plan certificates, explanations for common diagnostic tests and procedures, differential diagnoses, or even reference values. Moreover, most summaries of diseases are simply text blocks, with few hypertext or other links to relevant literature, topics or providers. The integration of forms and search engines is also fairly minimal, given such basic uses as: plan and address changes, personalized benefits inquiries, interactive maps, questionnaires, pilot tests, complaints, compliments, etc.

Professionals in many fields, not only health care, are frequently chastised for their failure to offer customers a clearly-defined role in specifying the product. The plumber says simply, "you need a new drain pipe." The physician states, "you need to start a beta blocker," not necessarily addressing the consequences of each alternative.

Basic concerns, such as clarifying terse statements about diagnosis or treatment, or why certain questions are asked during an examination, remain difficult to address under the hierarchial, parochial structure of most Web-based health information. Many sites restate information that can be found elsewhere, rather than linking to other sites as appropriate and building a site based on their individual strengths. Content providers benefit when the highest quality, most creatively-presented Web-based information on every conceivable health topic is available readily to their patients and other constituents.

Three overlapping viewpoints characterize Web-based health care information:

Model: The Web facilitates information flow, increases sophistication about health, and fosters realistic expectations and the management thereof. Over time, the Web and other media will successfully enhance self-responsibility, lifestyle changes, and health maintenance as opposed to symptom-focused, quick-fix care.

Marketing Tool: The Web should be utilized as any other marketing tool. Thus, content providers on the Web should focus on specific messages, complement other marketing tactics, and mediate unfavorable messages whenever possible. To do otherwise cedes the Web to a random assortment of uncontrolled data.

Menace: The Web is more a menace to appropriate care-seeking than a medium for education or marketing. It allows low-quality messages to be accessed repeatedly, thus continually lowering the aggregate quality of health care information available. "Quackery" assumes an artificial importance since Web users cannot differentiate sources of information and their credentials as easily as with other media. Thus, providers and other responsible parties should focus their efforts on traditional media in an attempt to counter and perhaps neutralize the Web's impact.

Certain characteristics of the Internet strongly influence the usage of health care information, such as:

Proliferation of unattributable and unchallenged assertions: A statement such as, "Mayo Clinic has approved this treatment," will often go unchallenged, because an appropriate Mayo Clinic official is unlikely to see it, and its originator may be difficult to identify, contact or hold responsible for accuracy.

Increased recipient expectations: The availability of "information on demand," sorted and classified various ways by search engines and subject trees, increases user expectations about the accuracy, completeness and timeliness of information found. Even blatantly-commercial sites sometimes appear more objective than they really are, due to a novelty factor and cross-links from more reputable sites.

Instant gratification for skilled information seekers: Within seconds on the Web, one can identify the chemical formula of a drug, specific indications and side effects, arguments for and against its use, and the results of recent clinical trials. A differential diagnosis, based on symptoms and a search engine, can be obtained in less time than it takes to dial a doctor's office and speak with a nurse.

Link between experience and usage: The extent and nature of users' knowledge of medicine and the Web affects the type of search conducted, which, in turn helps determine the degree to which information is verified and acted on, if at all. Content providers thus must consider the advantages of greater control over the search process. Except for limited-access professional societies, organizations do not generally group information by specificity or required sophistication. However, the speed and accuracy with which specific information is accessible largely depends on the expertise of the searcher, and the learning curve is steeper for those easily frustrated with "dead ends." Thus, the development and use of software which records mouse clicks, text selections, and pages viewed, and offers increasingly-appropriate selection options and questions to facilitate a user's exploration is particularly relevant to health information sites.

Ease of "broadband" transmission: The Web, IRC and Usenet allow a single individual to simultaneously transmit a message to thousands of potential recipients and to successively and unpredictably reach thousands more, who in turn may transmit information secondarily through "rebroadcast" and/or personal contact.

Meta-awareness of information: The Web, spanning ancient texts to "real-time" updates, encourages "meta-awareness" of health care and many other societal issues. That is, it stores and offers parallel information about both health care products and their existence as products. The ability to view simultaneously material designed for the lay public and medical professionals accelerates meta-awareness.

The Web's medical strands will ultimately be judged by their relationship to the people, procedures, policies and politics which affect, or seek to affect, users' health status or organization. A tighter weave between information about care and the care itself will maximize the Web's benefits.

Shame and War Revisited

Adding Semantic Markup to HTML

Philip Greenspun
Laboratory for Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology

Abstract

"HTML represents the worst of two worlds. We could have taken a formatting language and added hypertext anchors so that users had beautifully designed documents on their desktops. We could have developed a powerful document structure language so that browsers could automatically do intelligent things with Web documents. What we *have* got with HTML is ugly documents without formatting *or* structural information." I wrote that in August 1994. In the intervening 18 months, style sheets have substantially enhanced HTML's formatting capabilities, but no progress has been made on the structure problem. I propose a class-based semantic markup system compatible with existing browsers and HTTP servers.

Introduction

> "Owing to the neglect of our defences and the mishandling of the German problem in the last five years, we seem to be very near the bleak choice between War and Shame. My feeling is that we shall choose Shame, and then have War thrown in a little later, on even more adverse terms than at present."

Winston Churchill in a letter to Lord Moyne, 1938 [Gilbert 1991]

If you asked a naive user what the Web would do for them, they'd probably say "I could ask my computer to find me the cheapest pair of blue jeans being sold on the Internet and 10 seconds later, I'd be staring at a photo of the product and being asked to confirm the purchase. I'd see an announcement for a concert and click a button on my Web browser to add the date to my calendar; the information would get transferred automatically."

We computer scientists know that the Web doesn't actually work this way for naive users. Of course, armed with 20 years of Internet experience and the latest in equipment and software, we computer scientists go out into the Web and... *fall into exactly the same morass*. When we find a conference announcement, we can't click the mouse and watch entries show up in our electronic calendars. We will have to wait for computers to develop natural language understanding and common sense reasoning. That doesn't seem like such a long way off until one reflects that, given the ability to understand language and reason a bit, the computer could go to college for four years and come back capable of taking over our job.

Recently adopted HTML styles sheets offer us a glimmer of hope on the formatting front. It may yet be possilbe to render a novel readbly in HTML. However, style sheets can't fix all of HTML's formatting deficiencies and certainly don't accomplish anything on the semantic tagging front.

Fixing the formatting problem; frames are not the answer (but maybe style sheets are)

When a reader connects to a $100,000 Web server via a $10,000/month T3 line, his first thought is likely to be "Wow, this document looks almost as good as it would if had been hastily printed out from a simple word processor." His second thought is likely to be "Wow, this document looks *only* almost as good as it would if had been hastily printed out from a simple word processor."

Increased formatting capabilities are fundamentally beneficial. It is more efficient for one person to spend a few days formatting a document well than for 20 million users to each spend five minutes formatting a document badly. Yet the original Web model was the latter. Users would edit resource files on the Unix machines or dialog boxes on their Macintosh to choose the fonts, sizes, and colors that best suited their hardware and taste. Still, when they were all done, *Travels with Samantha* and the Bible ended up looking more or less the same.

The Netscape extensions ushered in a new era of professional document design, but it hasn't all been for the best, especially the round introduced with Netscape 2.0. HTML documents may have looked clunky back in 1993 but at least they all worked the same. Users knew that if they saw something black, they should read it. If they saw something gray, that would be the background. If they saw something blue, they should click on it. Unlike CD-ROMs, web sites did not have sui generis navigation tools or colors that took a few minutes to learn. Web sites had *user interface stability*, the same thing that made the Macintosh's pull-down menus so successful (because the print command was always in the same place, even if it was sometimes grayed-out).

Netscape 1.1 allowed publishers to play with the background, text, link, and visited link colors. Oftentimes, a graphic designer would note that most of the text on a page was hyperlinks and therefore just make all the text black. Alternatively, he or she would choose a funky color for a background and then three more funky colors for text, link, and visited link. Either way, users have no way of knowing what is a hyperlink and what isn't. Oftentimes, designers get bored and change these colors even within a site.

Very creative publishers managed to use the Netsacpe 1.1 extensions to create Web documents that looked like book or magazine pages. They did this by dropping in thousands of references to transparent GIFs, painful for them but even more painful for the non-Netscape-enhanced user.

Frames, introduced with Netscape 2.0, give the user the coldest plunge into unfamiliar user interface yet. The "Back" button no longer undoes the last mouse click; it exits the site altogether. The space bar no longer scrolls down; the user has to first click the mouse in the frame containing the scroll bar. Screen space, the user's most precious resource, is wasted with ads, navigation "aids" that he has never seen before, and other items extraneous to the requested document.

Thanks to all of these Netscape extensions, the Web abounds with multi-frame, multi-color, multi-interfaced sites. Unfortunately, it still isn't possible to format a novel readably. I'll use *The English Patient* [Ondaatje 1992] as an example. Although its narrative style is about as unconventional as you'd expect for a Booker Prize winner, it is formatted very typically for a modern novel.

Sections are introduced with a substantial amount of whitespace (3 cm), a large capital letter about twice the height of the normal font, and the first few words in small caps. Paragraphs are typically separated by their first line being indented about three characters. Chronological or thematic breaks are denoted by vertical whitespace between paragraphs, anywhere from one line's worth to a couple of centimeters. If the thematic break has been large, it gets a lot of whitespace and the first line of the next paragraph is not indented. If the thematic break is small, it gets only a line of whitespace and the first line of the next paragraph *is* indented. So the "author's intent" needs to be expressed with tags like `<small-thematic-break>`. The "designer's intent" needs to be expressed with equations like `small-thematic-break = one line of whitespace.`

Style sheets, officially adopted as a standard on March 5, 1996 by most browser makers, make this possible in almost the manner I've described. I asked Hakon W. Lie, one of the authors of the style sheet proposal, for the most tasteful way to format *The English Patient*. He came back with the following:

```
<STYLE>
  P { text-indent : 3em }
  P.stb { margin-top: 12pt }
  P.mtb { margin-top: 24pt; text-indent : 0em}
  P.ltb { margin-top: 36pt; text-indent : 0em}
</STYLE>

<P CLASS=stb>Sample of small thematic break
<P>just an ordinary paragraph
<P CLASS=mtb>Sample of medium thematic break
<P CLASS=ltb>Sample of large thematic break
```

The cascading style sheet proposal that was ultimately successful rejected the idea of new tags because a document marked up with such tags would not have been valid under the HTML document type definition (DTD).

Is the formatting problem solved? I begged for style sheets in my August 1994 paper and now we have them, much better thought-out and more powerful than I envisioned. The author/designer intent split is captured nicely. So what is left to do? Style sheets don't let one publish mathematics, figures with captions, or dozens of other things faciliated by old languages like LaTeX or newer systems like Microsoft Word.

We could just add hyperlinks to LaTeX. This is more or less what a group of people at Los Alamos National Labs did a few years ago. I don't think there are really sound intellectual arguments against this approach, but sentiment seems to be on the side of keeping HTML. If we are indeed stuck with HTML, though, perhaps there is a better way to extend it.

Our methodology for extending HTML seems to be the following

1. sit down with a few formatting languages in common use
2. argue about which are the most commonly needed commands
3. argue about whether Web browser programmers are really up to the task of writing the code that implements those commands

At this rate, it will be the year 2000 before HTML is really powerful enough for most people, by which time it may have been replaced with Adobe's Portable Document Format (PDF).

I'd like to suggest an alternative approach:

1. choose a set of 100 documents that represent the spectrum of things we'd like to see on the Web
2. come up with a language capable of expressing the author's intent in 98 of those documents
3. come up with a language capable of expressing the designer's intent in 98 of those documents
4. add to HTML the semantics of the languages developed in the preceding two steps

Fixing the Structure Problem

Can the same approach solve the structure problem? What if we locked a bunch of librarians and a handful of programmers in a room together and made them think up every possible semantic slot that any Web document could ever want to fill. They'd come out with a list of thousands of fields, each one appropriate to at least a small class of documents.

An obvious reason why this wouldn't work is that the committee could never think of all the useful fields. Five years from now, people are going to want to do new, different, and unenvisioned things with the Web and Web clients. Thus, a decentralized revision and extension mechanism is essential for a structure system to be useful.

A deeper reason why this wouldn't work is that nobody would be able to write parsers and user interfaces for it. If a user is developing a Web document, does he want to see a flat list of 10,000 fields and go through each one to decide which is relevant? If you are programming a parser to do something interesting with Web documents, do you want to deal with arbitrary combinations of 10,000 fields?

Malone's Work on Semistructured Messages

Back in the early 1980s, Tom Malone and his collaborators at MIT developed the Information Lens, a system for sharing information within an organization. He demonstrated how classifying messages into a kind-of hierarchy facilitated the development of user interfaces. Figure 1 shows one of Malone's example hierarchies [**** insert Figure 6 from Malone's paper]. For each message type, there is an associated list of fields, some of which are inherited from superclasses. Consider the class `meeting-announcement`. Fields such as `to`, `from`, `cc`, and `subject` are inherited from the base class `message`. Fields such as `meeting-place` are associated with the class `meeting-annoucement` itself.

Each message type also has an associated list of suggested types for a reply message. For example, the suggested reply type for `meeting-announcement` is `request-for-information`. Most importantly, the decomposition of message types into a kind-of hierarchy allows the automatic generation of helpful user interfaces. For example, once the system knows that the user is writing a `lens-meeting-announcement`, that determines which fields are offered for filling and what defaults are presented. Fields having to do with software bugs or New York Times articles are not presented and fields such as `place` and `time` may be helpfully defaulted with the usual room and time.

What did Malone's team learn from this?

- That a very wide range of messages could be processed automatically. It was convenient for users to fill in lots of fields so messages typically had enough structure to enable fairly sophisticated automatic processing.
- That by not forcing users to fill out every field and by allowing users to insert arbitrary text in some fields, unusual situations could be handled gracefully.
- That making message types explicit facilitated the development of rules for automated processing. For example, a few lines of code sufficed to delete every New York Times article whose `article date` was prior to `today`.

Adapting Malone's Work to the Web

Where do we put the fields?

First of all, if we are not to break current clients, we need a place to put fields in an HTML document such that they won't be user-visible. Fortunately, the HTML level 2 specification provides just such a place in the form of the `META` element. `META` tags go in the head of an HTML document and include information about the document as a whole. For example

```
<meta name="type" content="conference-announcement">
<meta name="conference-name" content="WebNet-96">
<meta name="conference-location-brief" content="San Francisco">
<meta name="conference-location-full" content="Holiday Inn Golden Gateway Hotel,
<meta name="conference-date-start" content="16 October 1996">
<meta name="conference-date-end" content="19 October 1996">
<meta name="conference-papers-deadline" content="15 March 1996">
<meta name="conference-camera-ready-copy-deadline" content="1 August 1996">
```

would be part of the description for our conference and provides enough information for entries to be made automatically in a user's calendar.

It might not be pretty. It might not be compact. But it will work without causing any HTML level 2 client to choke.

There are a few obvious objections to this mechanism. The most serious objection is that duplicate information must be maintained consistently in two places. For example, if the conference organizers decide to change the papers deadline from 15 March to 20 March, they'll have to make that change both in the `META` element in the `HEAD` and in some human-readable area of the `BODY`.

An obvious solution is to expose the field names and contents to the reader directly, as is typically done with electronic mail and as is done in [Malone 1987]. When Malone added semiformal structure to hypertext [Malone 1989], he opted to continue exposing field names directly to users. However, that is not in the spirit of the Web; stylistically, the best Web documents are supposed to read like ordinary text.

A better long-term solution is a smart editor for authors that presents a form full of the relevant fields for the document type and from those fields generates human-readable text in the `BODY` of the document. When the author changes a field, the text in the `BODY` changes automatically. Thus, no human is ordinarily relied upon to maintain duplicate data.

How do we maintain the document type hierarchy?

Malone unfortunately cannot give us any guidance for maintaining a type hierarchy over a wide area network. He envisioned a system restricted to one organization. His object-oriented approach can give us some inspiration, however. Malone reports that a small amount of user-level programming sufficed to turn his structure-augmented hypertext system into a rather nice argument maintenance tool, complete with user-interface for both display and input [Malone 1989].

Whatever mechanism we propose, therefore, had better allow for an organization to develop further specialized types that facilitate clever processing and presentation. At the same time, should one of these hyperspecialized documents be let loose on the wider Internet, it should carry some type information understandable to unsuspecting clients. Once mechanism for doing this is the inclusion of an extra type specification:

```
<meta name="type" content="lanl-acl-conference-announcement">
<meta name="most-specific-public-type" content="conference-announcement">
```

In this case, the Los Alamos National Laboratory's Advanced Computing Laboratory has concocted a highly specialized type of conference announcement that permits extensive automated processing by Web clients throughout Los Alamos. However, should someone at MIT be looking at the conference announcement, his Web client would fail to recognize the type `lanl-acl-conference-announcement` and look at the `most-specific-public-type` field. As `conference-announcement` is a superclass of `lanl-acl-conference-announcement`, all the things that the MIT user's client is accustomed to doing with conference announcements should work with this one.

Nonhierarchical inheritance (also known as "multiple inheritance") is also important so that duplicate type hierarchies are not spawned. For example, the fact that a document is restricted to a group or company might possibly apply to *any* type of document. Should there be two identical trees, one rooted at `basic-document` and the other at `basic-internal-document`? Then we might imagine documents for which there is an access charge. Now we just need four identical trees, rooted at `basic-free-document, basic-metered-document, basic-internal-free-document, basic-internal-metered-document`. There is a better way and it was demonstrated in the MIT Lisp Machine Flavor system (a Smalltalk-inspired object system grafted onto Lisp around 1978): mixins. Mixins are orthogonal classes that can be combined in any order and with any of the classes in the standard kind-of hierarchy. Here are some example mixin classes:

Class Name	Fields Contributed	Comments

`draft-mixin`	`draft-expected-completion-date`
	`draft-address-for-comments`
	`draft-version-number`
	`draft-previous-version-url`
		User Agent displays "****DRAFT****" prominently, offers to look up previous version and show change bars.

`restricted-mixin`	`restricted-current-release-terms`
	`restricted-person-authorized-to-release`
	`restricted-expected-release-date`
	`restricted-authorized-access-specification` (explains who can access, possibly a domain name or list of networks)
		HTTP server watches for documents whose type inherits from this class and only delivers them to authorized users; non-authorized users sent an explanation with the name of a person who could authorize release.

If there are N mixins recognized in the public type registry, we might have to have 2^N classes for every class in the old kind-of hierarchy. That's one for every possible subset of mixins, so we'd have classes like `travel-magazine`, `travel-magazine-restricted`, `travel-magazine-draft`, `travel-magazine-draft-restricted`, etc. This doesn't seem like a great improvement on the 2^N identical trees situation.

However, if we allow documents to specify multiple types

```
    <meta name="types" content="travel-magazine restricted-mixin draft-mixin">
```

and build the final composite type at runtime in the content editor, HTTP server, and Web user agent, then we need only have one hierarchy plus a collection of independent orthogonal mixins. This presents no problem for programmers using modern computer languages such as Smalltalk and Common Lisp. These allow new type definitions at run-time and have had multiple inheritance for over a decade. A program implemented in a language that has purely static types, e.g, C++ or Java, is going to need to include its own dynamic type system, built from scratch and not based on the underlying language's type system.

We established then that we need multiple inheritance and distributed extensibility. A standard Internet approach to distributed maintenance of a hierarchy is found in the Domain Name System (DNS), where authority for a zone is parcelled out and that authority includes the ability to parcel out subzones [Stevens 1994; Mockapetris 1987a ; Mockapetris 1987b].

DNS-style type definition service might seem like overkill initially and would result in delays for pioneer users of document types. Without a substantial local cache, document type queries would have to be sent across the Internet for practically every Web document viewed. An alternative would be to have documents include their type definition code at the top or reference a URL where such a definition might be found. This is how it is done with style sheets.

Regardless of how the hierarchy is maintained, developing the initial core taxonomy is a daunting task. The taxonomies developed by librarians are only a partial solution because they do not generally concern themselves with the sorts of ephemera that constitute the bulk of Internet traffic. If we don't get the core taxonomy right, we won't reap the benefits of useful standard software.

Conclusions

BEST COPY AVAILABLE

Measured against the yardstick of "how well does this Internet thing work?", HTML is an underachiever. It lacks sufficient structural and formatting tags to render many documents comprehensible much less aesthetic, even with the addition of style sheets. The META tag can be exploited to implement a document typing system. We need to develop a hierarchy of document types to facilitate implementation of programs that automatically process Web documents. This type system must support multiple inheritance. If we fail to develop some kind of semantic tagging system, computers will be unable to render us any useful assistance with Web documents until the dawn of Artifical Intelligence, i.e., natural language understanding and common sense reasoning.

References

Martin Gilbert 1991. *Churchill A Life* Henry Holt & Company, New York, page 595

Lie, H.W., Bos, B. 1996. "Cascading Style Sheets, W3C Working Draft" (http://www.w3.org/pub/WWW/TR/WD-css1.html)

Malone, Thomas W., Grant, Kenneth R., Lai, Jum-Yew, Rao, Ramana, and Rosenblitt, David 1987. "Semistructured Messages are Surprisingly Useful for Computer-Supported Coordination." *ACM Transactions on Office Information Systems*, **5**, 2, pp. 115-131.

Malone, Thomas W., Yu, Keh-Chaing, Lee, Jintae 1989. *What Good are Semistructured Objects? Adding Semiformal Structure to Hypertext.* Center for Coordination Science Technical Report #102. M.I.T. Sloan School of Management, Cambridge, MA

Mockapetris, P.V. 1987a. "Domain Names: Concepts and Facilities," RFC 1034

Mockapetris, P.V. 1987b. "Domain Names: Concepts and Facilities," RFC 1035

Ondaatje, Michael 1992. *The English Patient*. Vintage International, New York

Stevens, W. Richard 1994. *TCP/IP Illustrated, Volume 1: The Protocols.* Addison-Wesley, Reading, Massachusetts

Professional Development for Educators via the Web

Keith A. Hall, Professor
The Ohio State University
College of Education
29 West Woodruff Avenue RA122
Columbus, Ohio USA 43210
hall.25@osu.edu

William Gathergood, Clinical Educator
The Ohio State University
College of Education and Reynoldsburg High School
1579 Huxley Drive
Columbus, Ohio USA 43227
gathergood.1@osu.edu

Differing Cultures: Bringing Minds Together

The culture of the university teacher education professor is different from the culture of the K-12 classroom educator; the culture of the secondary classroom educator is different from the culture of the elementary classroom educator; the culture of the urban high school student is different from the culture of the suburban or rural high school student; and the culture of the American middle school student is different from the culture of the Russian, Japanese, Australian, or United Kingdom middle school student. Web collaborations among these groups narrow the culture gap and lead to increased understanding.

A collaboration, started in 1990 and known as the Technology in Education Professional Development School (TIEPDS), was initiated to bring university teacher education faculty and professional K-12 classroom educators together to enhance the pre-service preparation of teachers, to facilitate professional development of current classroom educators, and to conduct inquiry into the preparation and practices of teachers. It has blossomed from a local collaboration to a composite of national and international collaborations among and between university faculty, K-12 classroom educators, pre-service teachers, university graduate students, and K-12 students.

The programmatic collaborations have evolved from financial investment of the University and Reynoldsburg City Schools in appointing a K-12 classroom educator as Clinical Educator in the College of Education (with a 50% release time from his high school teaching responsibilities) and adjusting the responsibilities of a College of Education professor to serve as co-coordinators of the TIEPDS and to allow for intensive field work in local schools. Although the co-coordinators are usually able to meet weekly for planning, only one telephone call has taken place between the two in nearly six years! The majority of the planning, idea development, writing, and creative endeavor has taken place via e-mail, mailing lists, and FTP--several times per day and frequently seven days per week.

The TIEPDS Forum, a weekly meeting of K-12 educators and university graduate students, has been a major tool to acclimate the participants to the power of on-line collaboration and encourage its use--taking us from the comfort of face-to-face class meetings to meetings of the intellect on the Web and establishing comfort there.

The Co-coordinators of The Ohio State University TIEPDS will extract, synthesize, and collaboratively present findings from a broad array of TIEPDS projects--not as unique project reports but rather as aggregated findings gathered across projects which speak to basic constructs of collaboration on the Web.

Literary works have been used as the vehicle for much of the collaborative work which will be reported here--Shakespeare, Mark Twain, Harper Lee (***To Kill a Mockingbird***), ***The Great Gatsby***, and Chekow. And literary personification--a technique in which teachers and students assume the persona of authors and characters from literature to help students better understand the significance of literature--has been instrumental in engaging participants in the lives of the characters and authors. A teacher may become Mark Twain, guiding young students through their first reading of Huck Finn; students can ask questions of Hamlet while reading the play; or teachers may expand their horizons as they change their race, gender and ethnic background to fit a role. So an elderly northern male will come to see ***To Kill a Mockingbird*** from the perspective of the author, Harper Lee a young female Southerner, by answering students' questions via e-mail as Harper Lee.

Leveling the Playing Field: Equality Among Collaborators

The features of electronic communications--asynchronous, text only, and think-time before responding--all contribute to the depth and richness that can be achieved through collaboration. The collaboration is strengthened because the partners are not disadvantaged by distance, schedules, or rank--only the displayed words must be considered during the dialogue--facilitating an intellectual closeness through personal anonymity. We've found that the non-visual or non-pictorial nature of e-mail strengthens the focus on the message rather than the person and in many cases has strengthened the relationship between the individuals. In fact, just the phrase "E-mail to ..." which appears on most web pages creates the sense (and potential) of personal collaboration. Successive collaborations build deeper meaning and relationships in the work.

Web of Thought: Ways to Make the World Wide Web Educationally Relevant

Delwyn L. Harnisch
Department of Educational Psychology
University of Illinois at Urbana-Champaign, USA
harnisch@ux6.cso.uiuc.edu

Vince Gracia
Office of Minority Student Affairs
University of Illinois at Urbana-Champaign, USA
vgracia@ux7.cso.uiuc.edu

Matt Markus
Axon Software, USA

When browsing the World Wide Web one comes across many forces that could be viewed as detrimental to the educational process. For example, the Web often places a higher premium on pizzazz than content. Also, the non-linear nature of hypertext navigation can be potentially distracting to some learners. Yet, the World Wide Web's ability to allow ordinary computer users to easily unleash the power of the Internet's massive information archives has helped fuel the drive to place schools on-line. As more educational institutions begin to make the on-line transition, we must formulate methods for Web-based instruction that ensure the realization of the educational potential of this medium. To this end, we have constructed four example areas on the Del Harnisch (DLH) homepage that emphasize the essence of Web-based education. We discuss the motivational, educational, and technical aspects of these examples and hope that they contribute to an on-going discussion concerning Web-based instruction.

Collaboration on the Web

Today's educational system is modeled upon a classroom setting that is basically hierarchical in nature. The educational process is marked by *cooperation* between teacher and students in order to achieve various mandated educational goals. Until recently, a similar top-down organizational structure dominated the business world. Now, however, organizational structures are increasingly being flattened - cooperation has given way to *collaboration*. Collaboration allows groups to build knowledge synergistically where learning becomes a dynamic and an exciting activity. Yet, despite the benefits of collaboration, less than 2 percent of the United States' education budget is used to purchase collaborative technologies [Leebaert 1995]. One of the main objectives of the DLH homepage is to heighten awareness about the impact of collaborative technologies in the classroom without placing schools at financial risk.

The first example of Web collaboration on the DLH homepage is the Virtual Tutor applet. The Virtual Tutor was developed as an example of a collaborative Web technology that would allow students to interact live with a remote tutor. The Virtual Tutor was designed with two objectives in mind. The first objective was to illustrate to the educational community the benefits of on-line education in a non-passive way. The second objective was to grant learners access to distributed educational resources. For example, the Virtual Tutor makes it possible to run a highly interactive after school homework help page. Also, students can use the Virtual Tutor to discuss special assignments and problems with real scientists. Technically, the Virtual Tutor is a software package written entirely in Sun Microsystems' Java language and is based on a client/server architecture. When a learner enters the Virtual Tutor page the Virtual Tutor applet is served to the learner's machine. Once the applet has started running it sets up a white-board and a chat area on the learner's side and connects back to a special Java based server running on the host. On the server side a similar white-board and chat area are spawned thus completing an interactive link between the student and the teacher. By making the Virtual Tutor bytecode available for non-commercial use we hope to facilitate the transfer of cutting edge Web technology from the laboratory to the classroom.

Another place where the DLH home page attempts to use collaboration is in the Research area. Our Research area is targeted toward professionals who want to learn about Prof. Harnisch's current educational research activities. A Research area, though, can also be used by teachers to aide in peer evaluation of assignments. Although the World Wide Web is currently not a full interactive hypertext publishing system as described in [Drexler 1986], several technical patches exist to complement the power of the Web. One such technical extension is the utilization of backlinking. Backlinking works by finding external HTML pages that reference a particular paper on your machine. Thus, backlinking allows others to review published material and leave comments which others can then reference from the original work. Backlinking can also be used to help create a "living" bibliography. Such a bibliography not only shows the references you used to construct the paper, but also reveals who then later used your research as a reference. On the technical side, the documents found in the research area are Microsoft Word 6.0 documents which have been transformed to HTML via the free Word 6.0 Internet Assistant available from Microsoft. Charts and graphs found in this area were constructed in Microsoft Excel and converted into GIF files through the use of screen capture software. Finally, the back links were implemented by using Digital's Alta Vista search engine's "link:" command and a CGI script.

Making it Relevant

In [Negroponte 1995], Nicholas Negroponte relates several stories about children using technology to explore their environment. This "hard fun," as Negroponte refers to it, emphasizes the fact that often the biggest barrier to achieving understanding is when the subject matter lacks relevance to the learner's life. Two areas on the DLH homepage are aimed specifically at demonstrating ways that various topics can be made relevant to individuals through the use of technology.

The first area where relevance is cultivated is in the Survey area. It is here that one can find form-based surveys on a variety of subjects. The purpose of the Survey area is two-fold. First, the Survey area can enhance the relevance of class projects and stimulate cross-curriculum discussion. For example, in social studies a class can design a questionnaire pertaining to preferences in a presidential race. Later, in math class, the data collected from the Web can be statistically analyzed. The Survey area may also be used to issue feedback on the homepage, educational issues, and offer suggestions for future development.

The Presentation area of the DLH homepage is an area where Prof. Harnisch and his students place Microsoft PowerPoint presentations on educational topics from his courses and advanced seminars. Many teachers may find the idea of a Presentation area quite attractive. A Presentation area can serve as a place students can go to review past lectures. Also, by placing individual and class presentations on-line students have an added incentive to produce quality work. Technically, a Presentation area based on Microsoft PowerPoint presentations is not difficult to construct since the majority of work is actually done in Microsoft PowerPoint.

Future Directions

These examples hint at some of the exciting educational benefits the World Wide Web may offer to education. As the educational community heads into the twenty-first century it is our duty to embrace the new technology and employ it in the classroom setting. However, we must be ever vigilant in integrating technology with classwork so the Web becomes a valuable educational tool and not a meaningless classroom distraction.

References

[Drexler 1986] Drexler, E. (1986). *Engines of creation*. Garden City, N.Y.: Anchor Press/Doubleday.

[Leebaert 1995] Leebaert, D. (1995). News from the frontiers. In D. Leebaert (Ed.), *The future of* *software* (pp. 1-26). Cambridge, Mass.: The MIT Press.

[Negroponte 1995] Negroponte, N. (1995). *Being digital*. New York: Alfred A. Knopf.

On Teaching and Collaborative Development of Web-based Multimedia Knowledge-Based Systems

Amir Hekmatpour [1]
IBM PowerPC Design Center
11400 Burnet RD MD 9730
Austin, Texas 78758
amir@ibmoto.com

Irfan Dawood
The University of Texas at Austin
Dept. of Mechanical Engineering
Austin, Texas 78712
irfan@ccwf.cc.utexas.edu

Priya Kalyanasundaram
The University of Texas at Austin
Dept. of Mechanical Engineering
Austin, Texas 78712
priya@ccwf.cc.utexas.edu

1.0 Introduction:

There has been a steady increase in research and development of interactive multimedia information systems. Such systems are utilized in industry, education (edutainment) and commerce (infotainment). Useful information systems need to deal with a broad spectrum of expertise which could neither be obtained from a single expert, nor be modeled using a single technique. The successful realization of such systems depends on the ability to provide need-based information and real-time end-user decision support. Multimedia knowledge-based systems (MMKBS) have been gaining popularity in the fields dealing with mission critical applications, knowledge intensive tasks and heuristic decision support. Internet and world wide web on the other hand have opened a spectrum of possibilities for distributed hypermedia repositories, rapid access to customized information and cooperative problem solving. This paper reports on a project-based graduate level course[2] at The University of Texas at Austin[3] on the theory and principles of net-based MMKBS and tools and techniques for development and deployment of such systems.

The course combines traditional classroom lectures and web-based interaction and information dissemination to incrementally introduce and engage students with MMKBS, Web, project management and team-based application development [4]. The course methodology has proved to be an interesting and engaging method for teaching the technology, as well as providing the students with a platform for rapid prototyping of useful web-based MMKBS systems. This paper gives an overview of the course, the team-based project development process and a typical course project.

2.0 Course Description:

Most engineering and computer science students have been exposed to the notion of using Artificial Intelligence in engineering applications but have not had a chance to participate in real world problem solving projects. In addition, in most multimedia related courses, application development principles, techniques, and tools are hardly covered beyond desktop publishing, presentation generation, and computer graphics. This course was developed to teach fundamentals of MMKBS as well as provide hands-on training for developing web-based multimedia knowledge-based systems.

The major focus of this course is on techniques and tools for cooperative development of intelligent net-based multimedia diagnostic and training applications. A broad range of topics are covered to introduce students to the fundamentals of knowledge-based systems, multimedia information systems and net-based solution delivery, as well as the state-of-the-art tools and methodologies for practical development and deployment of such systems [5].

3.0 Course Projects: [6]

The course emphasizes and encourages hands-on learning. So, in addition to lectures, homeworks and tests students are required to define, develop, test and deploy a net-based MMKBS of their choice. The class is split into teams of 3-4 students and each team is asked to develop a business plan for a MMKBS application of their choice. The business plan should include detailed functional description as well as business justification and development plan and schedules. Each team presents their business plan to the class and then exchange their business plan with another team for evaluation and critique.

Once the proposed projects are accepted, each team partitions their project according to the principles of object-oriented programming and modular system design. Each student is assigned a module, and is responsible for developing and testing the module according to the specifications. Lectures, reading materials, and assignments are tailored to enhance the progress of class projects. As new topics are covered in class, students add new functions or capabilities to their modules. Subjects presented in the class are arranged in the order necessary for such an incremental and hands-on training and development process. In the last part of the course, all modules are put together for final system prototype and test.

In this paper we describe one of the projects developed by a group of students taking the MMKBS course in Spring of 1996.

3.1 FixCarAC Project Development:

The focus of this project was to develop a net-based multimedia expert system to diagnose and repair automobile air conditioning systems [7]. The main goal was to practice and implement principles and topics presented in lectures. The secondary goal was to develop a useful system which can be made available to the public. The system's architecture was to allow incremental upgrade of knowledge base and the multimedia repository with minimal impact on its utilization and GUI interface.

This system was initially developed as a pilot project on one car only to test its ease of use, and adaptability to other cars. The initial phase of the system development focused on knowledge acquisition by interviewing A/C mechanics on the workings of a typical automobile A/C system. In addition, detailed interviews were also conducted to obtain specific information about the A/C system of the Honda Civic.

At the present stage, the system can be run in its entirety from the internet browser, or from the Nexpert Object shell with access to an internet browser. The two systems share many of the data files. If a user does not have access to a net browser, the Nexpert-based system could be adapted to run on its own stand-alone multimedia modules. The only change would be that the data files would need to be reconfigured to a different file format, and the pointers in the Nexpert code will need to be modified to point to the new data files.

4.0 References:

- [1] Amir Hekmatpour is also an Adjunct Faculty at the University of Texas at Austin, Dept. of Electrical and Computer Engineering (amir@ece.utexas.edu).
- [2] Course Home page: http://www.ece.utexas.edu/~amir/ut_mmkbs.html
- [3] The University of Texas at Austin: http://www.utexas.edu
- [4] Course Description: http://www.ece.utexas.edu/~amir/html/ut_desc.html
- [5] Course Topics: http://www.ece.utexas.edu/~amir/html/ut_notes.html
- [6] Course Projects: http://www.ece.utexas.edu/~amir/html/ut_notes.html
- [7] FixCarAC Project Home Page: http://www.ece.utexas.edu/~amir/sp96grp1

Distance Education on Demand: (Serving MPEG on the Internet)

Mark D. Hendricks
Electronic Information Specialist
IANR Communications and Information Technology
University of Nebraska-Lincoln
United States of America
e-mail: ianr023@unlvm.unl.edu
websites: citv.unl.edu, citvsgi1.unl.edu

I.- History

In the late summer of 1995, the Institute of Agriculture and Natural Resources (IANR) Communications and Information Technology (CIT) at the University of Nebraska - Lincoln (UNL), began to research and assemble a server for the purpose of experimenting with the possibility of distributing audio and video information via Internet. IANR has a long history of using distance education and information delivery, and this server was a logical exploratory step for IANR to take.

Putting the components together was my task, and I began by installing Linux on a 60 MHz Pentium Computer, and the Xing server software. I had a thirty day trial with the Xing software and had never seen Linux. After about a month, the system was up and running and everything seemed to work. We digitized some sample pieces and were able to send them around campus.

II. - Background Technology

MPEG, and specifically MPEG-1 for these purposes, is a standard for distributing audio and video established by the Motion Photographers Experts Group. "Streaming" is a process of transmitting data (including MPEG) over a transmission media (in this case, the Internet), without having to copy the data to a storage device before viewing the data. What this means in that when you are logged on to a server that can "stream" data, you do not have to wait and download a file before you can use it. Streaming takes on an even greater significance when you consider the size and nature of the MPEG data we have set out to distribute. The MPEG files can easily exceed 100 MB and would take a long time to view if you had to download the file before you could view at it.

The Xing software we use requires some description. It consists of several software and hardware components. The server software I have discussed early is a MPEG database server. The browser software is a software only MPEG decoder, and can be used to view video over bandwidth as low as 14.4 bps, on PC's, Mac's and Unix machines. The server software is capable of transmitting MPEG streams at variable rates, and can actually take an MPEG stream from an audio or video encoder and simultaneously transmit the signal at several different bandwidths. For example I could serve a live feed from a video conference at 12 bps, 24 bps, 52 bps and 112 bps, so that people using 14.4 and 28.8 bps modems, 56 K lines and ISDN connections could view the conference at the best bandwidth their systems could support. MPEG and LBR files can also be stored on the hard drive for on demand replay and can be served at variable rates.

III. - Testing

By mid-October 1995, I had the server up and had received and configured the audio encoder. I was running NCSA's HTTPD server software and developed some web pages which incorporated a cgi call to the Xing server software. These allowed people using web browsers to select icons on web pages to receive MPEG and LBR data. The hardware and software was up, the interface was in place, and everything was working, so we decided we needed to know how many people we could distribute information to simultaneously.

Now, if you live in Lincoln, Nebraska, and you work for the University, and you need some content to get a bunch of people to test, one subject surfaces above all as common ground to encourage as much participation as possible -- football! So we met with Bill Byrne, UNL Athletic Director, and Chuck Jewell, General Manager of KFAB Radio, and received permission to do a live broadcast (KFAB radio audio) of the last three Nebraska home games. I received e-mail from Australia, Asia, and Europe as well as from all over the US asking me about software, saying how "cool" this was, talking about how they were displaced huskers, etc.

But the tests were game days, when three to five hundred people would log into http://citv.unl.edu and listen to the game. This data was served over a 4 mps token ring card from a Pentium 60 computer. From the nature of my mail, I would guess most people were able to get logged on and listen to the game.

IV. - Current Implementations

Beginning the first week of January, 1996, CIT started distributing six four and one-half audio news packages weekly. The content of these packages are time sensitive news items produced by CIT for release to Nebraska radio stations, and contain reports from specialists from all areas of IANR. This site is located at http://citv.unl.edu/almanac.htm, and the packages are updated every Friday.

The audio is digitized to .wav file format and converted to LBR format using a software product from Xing. This takes about 35 minutes, because our system has to digitize the audio in real time. After converting the data, it is moved to the server and entered in the servers database. The database is then recompiled. Two web pages are updated and the server is restarted. The whole process takes about an hour. One other note, the .wav files start out at about 5-6 MB and after conversion to LBR format are around 250-300K in size.

V. - The Near Future

Other projects on the horizon will probably include integrating web pages with content containing video and/or audio clips programs, as well as links to other related on-line information. This is timely information delivered to meet the needs of Nebraskans, based on incoming calls, and e-mail questions. One project, called EduPort, involves high-quality digitization of video and animation, including archives from Nebraska and the National Film Archive. Much of this material, due to it's fragile and/or costly nature, is not easily accessible by the general public, let alone teachers and other educators.

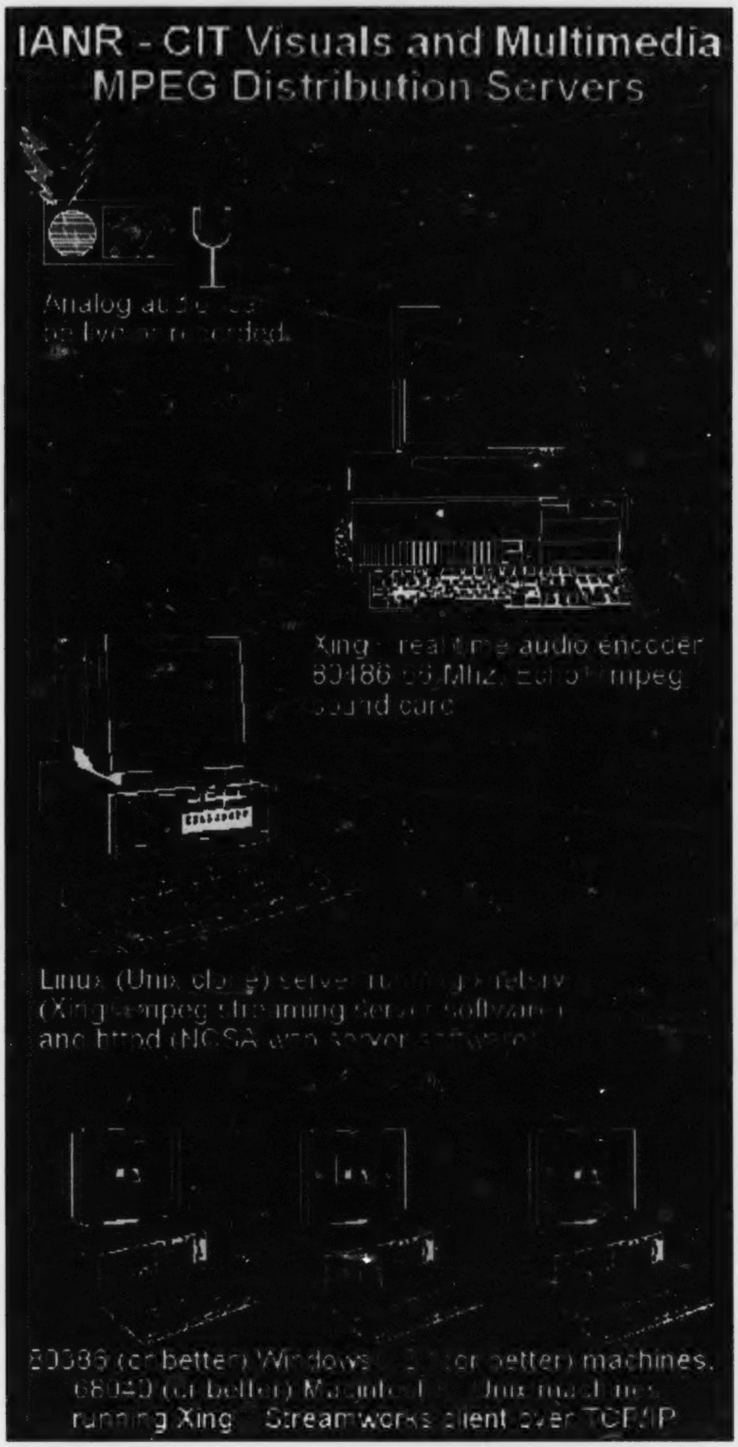
VI. - Implications of the Technology

With the infrastructure expanding and cost of this technology falling, I would submit that the application of these technologies will continue to expand and that the pace of installation will continue to excel, the economic and technological friction is rapidly decreasing. These are reasons enough for the technology to succeed.

But there is much more to this picture, education on demand means I can watch my professors 2 p.m. lecture again at 8 p.m. and even 2 a.m., and that I can watch it over and over, and skip to the parts I need to review. It means I can also review last weeks lecture, and perhaps other related material grouped with this information. Instructors can continue to update their information in real time, and point to other recommended resources, both on-line and off-line. Resources can be used 24 hours a day, on every continent, whenever the user wants to access the information.

VII. - Conclusion

If we consider where MPEG streaming technology is at today and compare it for instance with the aircraft industry, it is 1903 and Kitty Hawk is barely behind us. It will be 66 years before we will be on the moon, but things are starting to happen. I have only briefly discussed the implications of MPEG streaming in the near future, but at the current rate of change, I do not feel that I can speculate further, other that stating that it will get better, faster and more usable as the technology and delivery systems mature.


IANR - CIT Visuals and Multimedia
MPEG Distribution Servers
Analog audio can be live or recorded.
Xing real time audio encoder 80486 66 Mhz, Echo mpeg sound card
Linux (Unix clone) server running ... (Xing mpeg streaming server software) and httpd (NCSA web server software)
80386 (or better) Windows (or better) machines, 68040 (or better) Macintosh Unix machines running Xing Streamworks client over TCP/IP

BEST COPY AVAILABLE

Multi-level Navigation of a Document Space

Misook Heo
Emile L. Morse
Swantje Willms
Michael B. Spring

email: mxhst,elm2,swan,spring@sis.pitt.edu
http://www.lis.pitt.edu/~elm2/WebNet96.html

Department of Information Science and Telecommunications
University of Pittsburgh
Pittsburgh, PA 15260 USA

Navigation of spaces with more than two dimensions via a two dimensional display using a mouse is difficult, even when the space and the objects depicted are real-world objects. Navigation becomes more difficult when the space and the objects have no or minimal natural mappings to physical spaces. While we believe the mechanisms and approach described here are applicable to any situation in which abstract data is mapped to a physical display, this particular implementation is explored within the context of CASCADE (Computer Augmented Support for Collaborative Authoring and Document Editing), a research testbed for exploring the development of agents and tools to support collaborative authoring of documents [Morse et al. 1996]. Building on the work of [Spring and Jennings 1993] related to rules for mapping abstract data to virtual spaces, this research explores the use of a set of interrelated tools to assist a user in navigation of document collections. The research examines the navigational tools in isolation and in combinations.

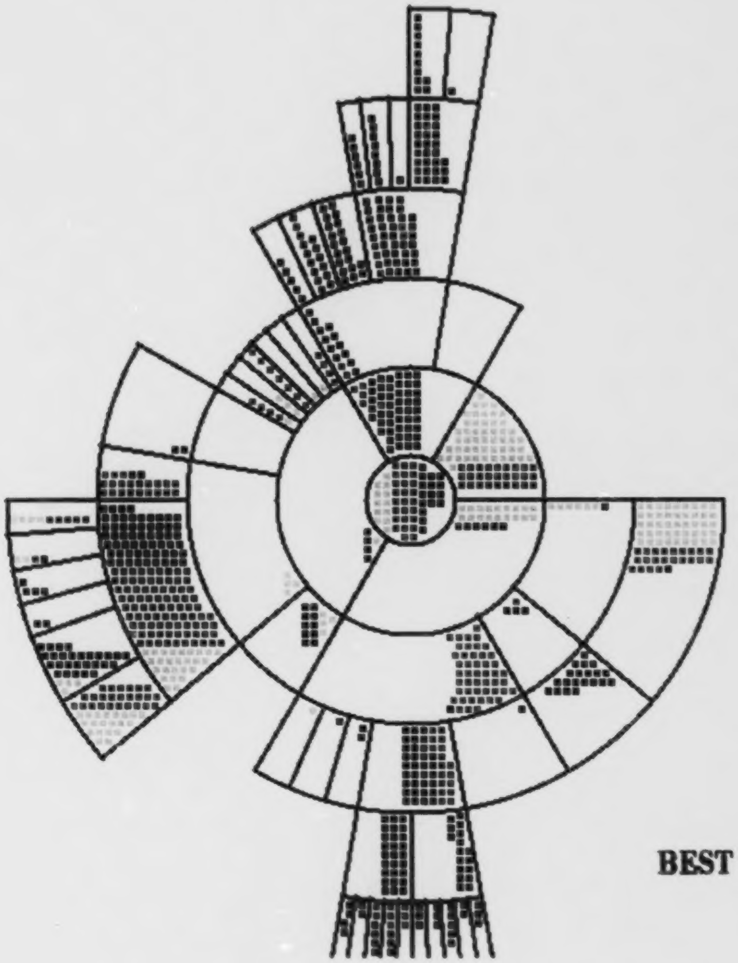
The tools may be thought of as a set of maps at different scales which can be used together to aid movement through complex, abstract document spaces. We begin with the assumption that documents and document components are objects, these objects are related and have attributes. Each tool looks at one or more of these features. The tools currently implemented for CASCADE include:

- **Docuverse.** This view depicts objects and selected attributes. It provides an overview of the document space and may highlight broad areas of interest based on textures that may develop from depiction of selected attributes such as size, age or other user-selectable document metadata. The docuverse view shows minimal structure, but depicts general structural features via proximity of objects. Selecting an area of the docuverse would instantiate the Webview.
- **Webview.** The Webview is a traditional tree browser showing typed relationships between objects and selected object attributes. Individual documents and links between documents are viewed on a large pannable canvas. While the prototype currently displays only hierarchical links, future versions will implement mechanisms for displaying other types of links as well. Selecting a container object in the tree causes the root of the tree to change -- to the new container. Selecting a leaf node -- a document object, causes the document object to be displayed, making Landmark tools accessible.
- **Landmarks.** Given a single document, intra-document tools are used to aid navigation. Landmarks within a document may be chosen to reflect the semantic, structural or other attributes of the document. In the current implementation, two landmark maps may be appended to a document in CASCADE. Each provides a view of selected attributes as seen in the displayed part of the document as well as the parts of the document not displayed.
- **Preview.** Preview provides the reader with previews of the destinations that can be reached from the current location.

Docuverse

Docuverse represents a large document space by selecting only one or two attributes of the documents in the space. [Shneiderman 1991] uses a similar technique known as Treemaps which renders both files and directories as bars. It is sometimes difficult to distinguish files from directories when complex document spaces are depicted in Treemaps. In contrast, the Docuverse depicts files as colored dots and directories as enclosing boxes, so that even with a complex document space, differentiation of files from directories is intuitive. Figure 1 shows a 6-level directory hierarchy which contains 1000 files; the rendering was made in a 600 x 600 pixel space. Documents have been represented using a 5 X 5 block of pixels (16 pixels for the file and one horizontal and one vertical edge for a border). Theoretically, about 10,000 documents could be rendered in this space. In reality, the number of directories and the density of files in those directories affects the upper limit of files that may be displayed. In the figure color has been mapped to age with yellow meaning the document is new and brown meaning that it is old.

We anticipate that Docuverse views, because of the time required for the system to navigate the document space to produce them, will be composed periodically and saved as bitmaps that might be rapidly rendered by the system. Experimental tests will determine what characteristics, e.g., size, age, or type, are of most interest to users. We also anticipate that users may wish to use patterns of change in maps as a focus mechanism. Thus, sequences of maps taken over time might be played like a movie to show changes in activity across broad areas of the Docuverse.



BEST COPY AVAILABLE

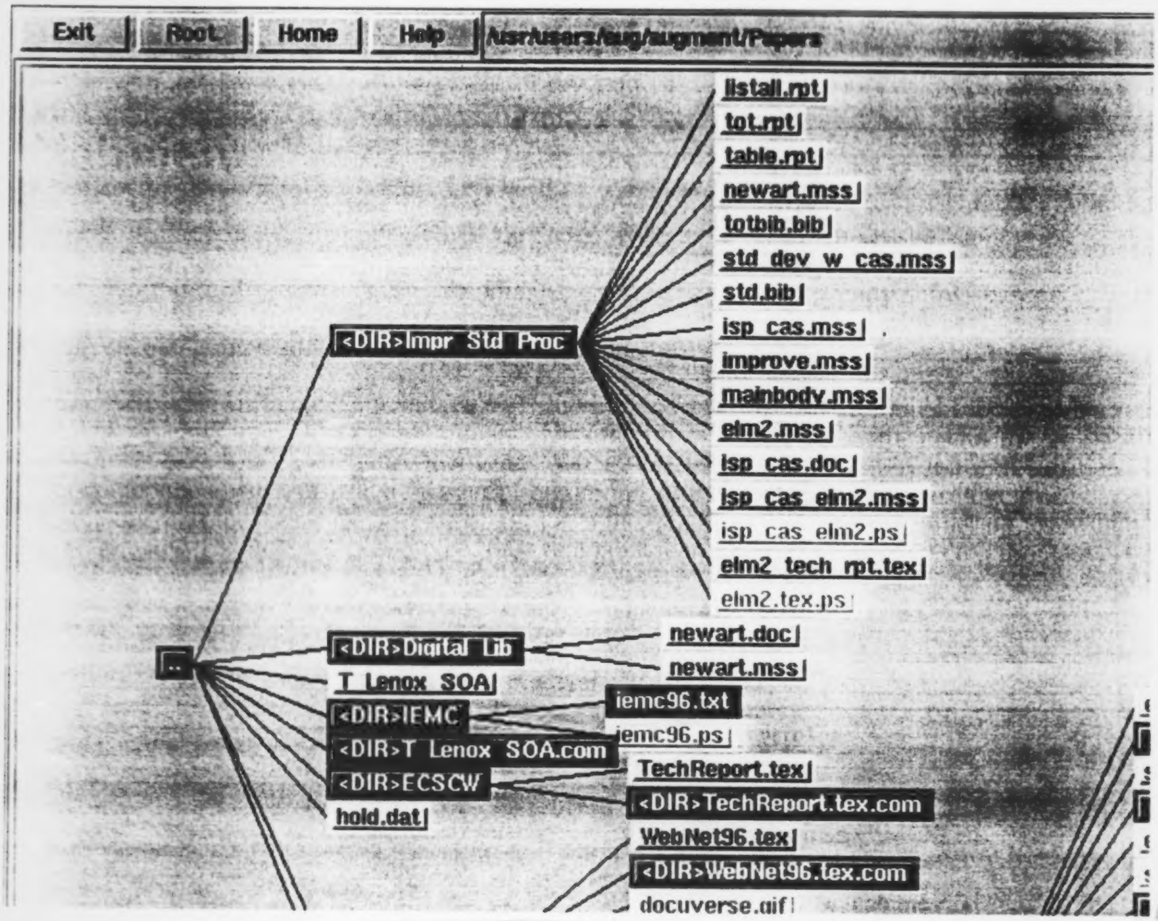
Figure 1: Docuverse View

The Docuverse shows selected characteristics of the objects in a broad expanse of the document space. For more detailed views, the user would shift his focus to the Webview browser.

Webview

Webview provides a tree-based structural display of part of the document. It shows individual document components and the links between them. Unlike the Docuverse, in addition to structure it displays multiple attributes of the components. The local vicinity currently shows parents, siblings, and descendents. Webview will also show links to nodes that are far away, e.g., cross references to documents in a distant area of the document space.

Figure 2 shows a four-level subtree. Color is used in this view to show component type -- e.g., comment, objection, graphic, etc. Color can alternatively be used to render the age of the document as was the case with the Docuverse. We are currently exploring how multiple attributes might be depicted using shape, color, saturation, texture etc. The displayed tree is developed dynamically, and the user may traverse up or down the tree by clicking on container nodes, which become the new root node. Beyond type, the node label identifies the owner and creation date or title of the node.


Exit
Root
Home
Help
/usr/users/aug/augment/Papers
listall.rpt
tot.rpt
table.rpt
newart.mss
totbib.bib
std dev w cas.mss
std.bib
isp cas.mss
improve.mss
mainbody.mss
elm2.mss
isp cas.doc
isp cas elm2.mss
isp cas elm2.ps
elm2 tech rpt.tex
elm2.tex.ps
<DIR>Impr Std Proc
..
<DIR>Digital Lib
newart.doc
newart.mss
T Lenox SOA
<DIR>IEMC
iemc96.txt
iemc96.ps
<DIR>T Lenox SOA.com
<DIR>ECSCW
TechReport.tex
<DIR>TechReport.tex.com
hold.dat
WebNet96.tex
<DIR>WebNet96.tex.com
docuverse.gif

BEST COPY AVAILABLE

Figure 2: Webview

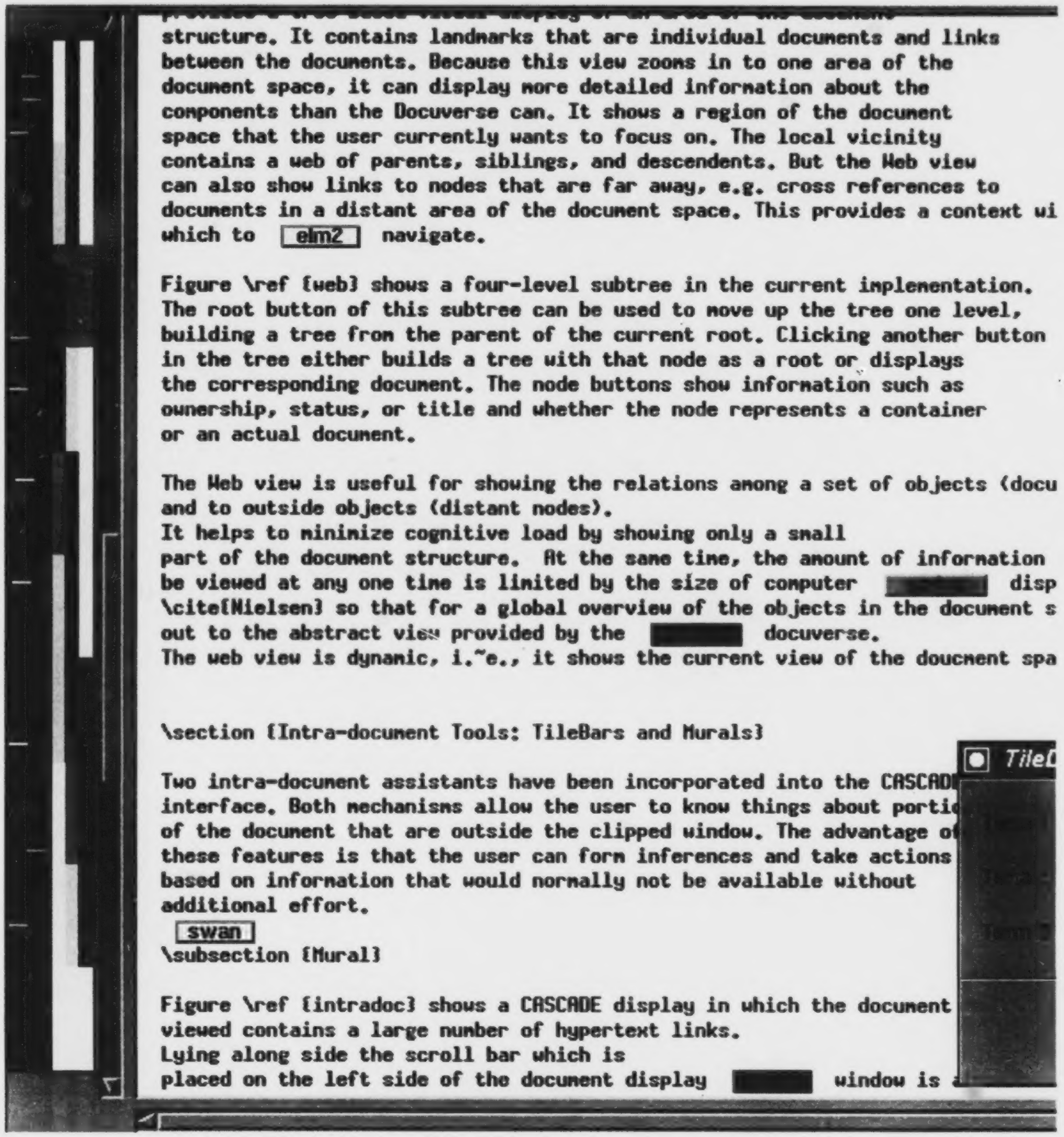
The Webview is useful for showing the relations among a set of objects (documents). It helps to minimize cognitive load by showing only a small part of the document structure.

Landmarks

Two Landmark tools have been incorporated into the CASCADE interface. They allow the user to know things about the overall document including portions not visible in the display window allowing the user to form inferences and take actions based on this information. The two Landmark tools are Mural and Tilebars. Tilebars is based on the work of Hearst et al. [Hearst 1995].

Mural

Figure 3 shows a CASCADE display in which the document which is being viewed contains a large number of hypertext links.



structure. It contains landmarks that are individual documents and links
between the documents. Because this view zooms in to one area of the
document space, it can display more detailed information about the
components than the Docuverse can. It shows a region of the document
space that the user currently wants to focus on. The local vicinity
contains a web of parents, siblings, and descendents. But the Web view
can also show links to nodes that are far away, e.g. cross references to
documents in a distant area of the document space. This provides a context wi
which to elm2 navigate.

Figure \ref {web} shows a four-level subtree in the current implementation.
The root button of this subtree can be used to move up the tree one level,
building a tree from the parent of the current root. Clicking another button
in the tree either builds a tree with that node as a root or displays
the corresponding document. The node buttons show information such as
ownership, status, or title and whether the node represents a container
or an actual document.

The Web view is useful for showing the relations among a set of objects (docu
and to outside objects (distant nodes).
It helps to minimize cognitive load by showing only a small
part of the document structure. At the same time, the amount of information
be viewed at any one time is limited by the size of computer [redacted] disp
\cite{Nielsen} so that for a global overview of the objects in the document s
out to the abstract view provided by the [redacted] docuverse.
The web view is dynamic, i.~e., it shows the current view of the doucment spa

\section {Intra-document Tools: TileBars and Murals}

Two intra-document assistants have been incorporated into the CASCADE
interface. Both mechanisms allow the user to know things about portio
of the document that are outside the clipped window. The advantage of
these features is that the user can form inferences and take actions
based on information that would normally not be available without
additional effort.
swan
\subsection {Mural}

Figure \ref {intradoc} shows a CASCADE display in which the document
viewed contains a large number of hypertext links.
Lying along side the scroll bar which is
placed on the left side of the document display [redacted] window is

Figure 3: Intradocument View showing both Mural and TileBar; Link Preview

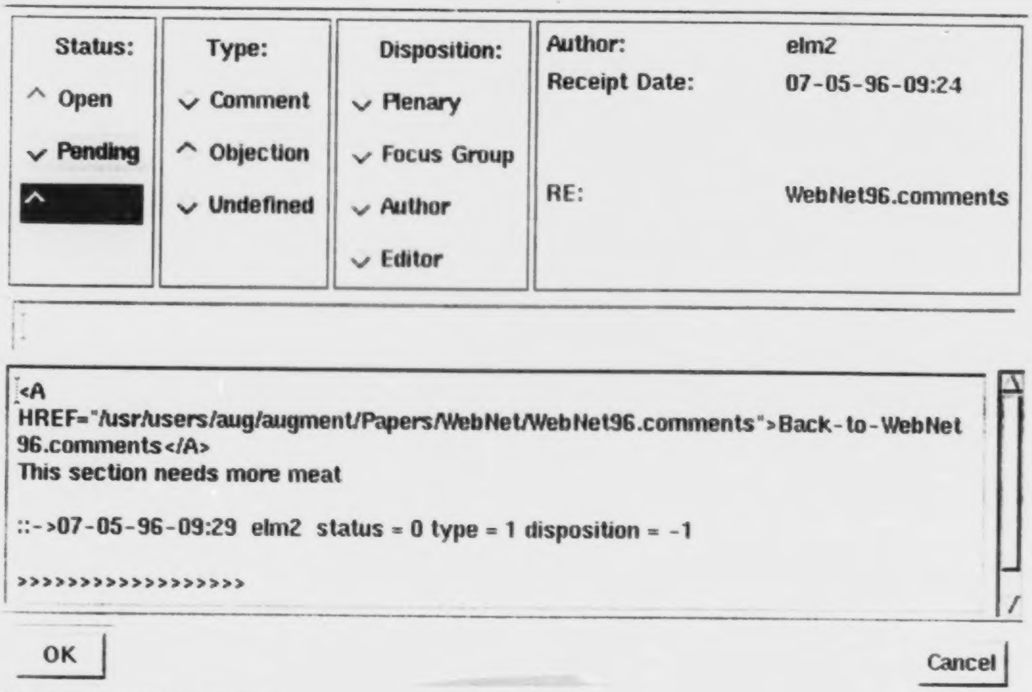
The Mural is placed to the far left of the scroll bar of the document display window. It represents a miniaturized view of the entire document. The Mural panel represents the full document. The portion of the panel co-extensive with the scroll bar thumb represents the portion of the document visible in the display window. Each of the color-coded rectangles in the Mural indicates the location and type of a link in the document. Using the Mural, the user can immediately see all the links in the document and move to any part of the document that he/she finds interesting. The Mural may also depict other structural features of a document such as structural copymarks or versioning information.

TileBar

BEST COPY AVAILABLE

The panel between the Mural and the scroll bar in Figure 3 is a TileBar. It uses the same conventions as the Mural in that it represents the entire document and the portion co-extensive with the scroll bar thumb represents the visible portion of the document. The TileBar is sectioned based on a semantic analysis of the document. The user then selects one, two or three groups of terms and the density of occurrences in each section is shown -- white indicating the absence of the term and black indicating a high occurrence. The figure shows that Term set 1 occurs most frequently in the first third of the document while the terms in the third set start to be found frequently in the section of the document currently being viewed.

Preview



Author: elm2
Receipt Date: 07-05-96-09:24
RE: WebNet96.comments

```
<A
HREF="/usr/users/aug/augment/Papers/WebNet/WebNet96.comments">Back-to-WebNet
96.comments</A>
This section needs more meat

::->07-05-96-09:29  elm2  status = 0 type = 1 disposition = -1

>>>>>>>>>>>>>>>>>>
```

Figure 4: Link Preview

Many research projects have used some kind of lens or link preview to provide additional information about a link before it is traversed. One of our systems, [Olsen et al. 1993] provided a lens with user-selectable contents. Preview represents the relationship between the current document and a linked document by showing metadata about the linked document within the document currently being viewed. The Preview has two types of views. One is an explicit view which displays only abstract ideas about the linked document such as title, author, or status. This view is shown in Figure 3 as the colored buttons indicating the location of hypertext links. From this view the reader can know who the author of the comment is (label information) as well as the comment type (color coding as in the Webview). The advantage of the explicit view is that it shows the information about all the visible links at one time. The second view is a popup lens which shows more specific information such as date of creation, length of the document, abstract and so on (Figure 4). The popup lens view shows a more detailed preview. Multiple popup lenses can be viewed simultaneously but their larger size places limits on how many can be used effectively.

Conclusion

Navigation through large document spaces, and abstract spaces in general, requires a number of specialized tools. Based on the assumption that a document space contains both objects and relations, we have developed a set of interrelated tools that are currently being tested and refined using the CASCADE testbed. Currently four tools -- Docuverse, Webview, Landmarks, and Preview -- are being used. It appears that each tool will play a role at different times in the navigation process. The tools have been linked such that appropriate actions in one tool invoke one of the other tools allowing the user to zoom in and out, gather detailed data as needed, and preview actions.

As indicated above, this work is still in a formative stage of development. Plans for future work include finding the optimal spatial representation for the Docuverse, exploring the use of time lapse representations of the Docuverse, looking at perspective representations of 2D and 3D trees to depict more information in the Webview, exploring the limits of data overloading in the Webview for expert and novice users, using Mural to render implicit structure as well as explicit structure.

References

[Hearst 1995] Hearst, M.A. (1995). TileBars: Visualization of term distribution information in full text information access. *Human Factors in Computing Systems, CHI '95 Conference Proceedings*, Denver, CO, May 7-11, 1995, 59-66.
[Morse etl. 1996] Morse, E.L., Willms, S., Pollak, W., Meyers, C. and Spring, M.B. (1996). *The Impact of Collaborative Authoring Software on the Standardization Process*, SIS Research Reports, LIS061/IS96002, Department of Information Science and Telecommunications, University of Pittsburgh.
[Olsen et al. 1993] Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B., and Williams, J.G. (1993). Visualization of a Document Collection with Implicit and Explicit Links: The VIBE System. *Scandinavian Journal of Information Systems*, 5:79-95.
[Shneiderman 1991] Shneiderman, B. (1991). Treemaps: a space-filling approach to the visualization of hierarchical information structures, In *Proceedings of the 2nd International IEEE Visualization Conference*, San Diego, October 1991, 284-291.
[Spring & Jennings 1993] Spring, M.B. and Jennings, M.C. (1993). Virtual Reality and Abstract Data: Virtualizing Information. *Virtual Reality World*, Spring 1993, 1(1):c-m.

Using a WWW Server to Access and Manage a Gemstone Smalltalk Server Supporting a Virtual Collaborative Learning Organisation

Frédéric HOOGSTOEL
E-mail : Frederic.Hoogstoel@univ-lille1.fr

Grégory BOURGUIN
E-mail : Gregory.Bourguin@univ-lille1.fr

Laboratoire TRIGONE Bâtiment B6
Université des Sciences et Technologies de Lille
59655 VILLENEUVE D'ASCQ
FRANCE.

I. Introduction

The Co-Learn Web Interface is the first step of the integration of a full CSCL organisation support system with the World Wide Web. It allows users to browse and manage the co-operation settings of an Open and Distance Learning Environment according to their roles in the virtual organisation. After the first part which describes the Co-Learn background, the second part sets out the Co-Learn Web Interface and its mechanisms. The last part concludes by showing the limits of our first implementation and presenting the perspectives of extensions of this work.

II. Co-Learn Background.

The Co-Learn environment has been designed to support Computer Supported Collaborative Learning in training institutions. It is a kind of global CSCW environment, like studied by the MOCCA project and described by [Benford 1991]. This means that Co-Learn integrates long-term/short-term and asynchronous/real-time collaboration tools. It allows the management of persistent collaborative environments represented by the Virtual Room metaphor [Madsen 1989][Hammainen, Condon 1991]. The goal of Co-Learn is to support the six issues we have identified for CSCW in organisations [Hoogstoel 1995] : communication, co-ordination, implication, cohesion, organisation and information sharing. The key issue of such an environment is tailorability. The Virtual Room is an elementary unit of the Organisation Space and a macroscopic element of the Activities Space (similarly to a *workspace* in CoopWWW [Appelt 1996] or a *locale* in wOrlds [Fitzpatrick, Kaplan, Tolone 1995]. The functional tailorability is assumed by allowing organisation members to instance and tailor Virtual Rooms. To tailor a Virtual Room, you choose the activities, documents and participants to include and you assign the roles of the participants in the activities. Since managing such a tailorable environment is very complex, the system proposes default configurations that can be easily modified. By offering basic tools and collaboration policies and allowing the creation of new ones, the system aims to constitute a meta-environment supporting the bootstrapping strategy, like envisioned by [Engelbart 1992].

III. The Co-Learn Web Interface.

1. Goals of the Co-Learn Web Environment.

Strong reasons of choosing the Web as an environment for co-operative work have been broadly discussed. For example, the last *ERCIM workshop on CSCW and the Web* has recalled that Web supports standardised hypermedia data handling, global address space, large scale and distribution, offers a well-known interface and constitutes an open system technology [Walther 1996] [Dix 1996].

Actually, we can notice a growth of the number of various groupware tools integrated into pages on the WWW. But these tools are usually not integrated in a Co-Learn-like global CSCW environment. The role notion is generally not supported, and when it is, there is no sharing or coherent managing of the user roles between the tools. On the other side, the Co-Learn CSCW environment would benefit from its integration in WWW environment because of the WWW advantages exposed before. That's why we are currently porting Co-Learn on WWW. This paper describes the first completed step of this work which concerns the management of the co-operation settings. This new management environment allows users to manage the organisation from any Web Browser as it was only possible before with the Windows Co-Learn administration tool. Moreover this environment can be used as an Animation Assistant [Hoogstoel 1992] during the co-operation sessions in virtual rooms.

2. Functionality Offered To Users.

The management functions can be classified by two aspects of the user profile : his expertise level and his roles. The *basic management facilities* of Co-Learn allow to tune the organisational and structural context of co-operation. It handles organisational and structural roles, courses, virtual rooms and users. The *advanced management facilities* of Co-Learn concern the fine-grained set-up of the co-operation inside the virtual rooms : it allows to choose tools, functional roles and collaboration modes. The *meta-management facilities* of Co-Learn support bootstrapping by defining new subclasses of co-operation objects (rooms, roles, co-operation modes). A set of rules defines permissions of users to do these actions, according to their roles. The *Users Administration* role is necessary to handle users at the level of the organisation. The *Structure Administration* role allows to perform the basic and advanced management functions handling the courses and the rooms. A *Teacher* in a room can perform basic and advanced management functions modifying the objects inside the room.

3. Design and Realisation.

a) Client-Server Communication Model.

The initial requirement was to move as it was the Windows Co-Learn administration tool on WWW. The existing interface was based on Windows forms and the administration tool didn't provide group feedback. Contrary to most of the Co-Learn tools which rely on a peer to peer communication model (*which would require to be implemented in a Java-like language*), the administration tool relies on a usual client-server communication model well supported by HTTP and the submit mechanism. We chose Perl to write CGI scripts because of its portability and string treatment facilities.

b) Data Handling and Dynamic View Generation.

As described before, the *Co-Learn web interface* allows User to inspect, create or modify data of the organisation model according to his role in the organisation (*see the User Model*). This data can be accessed through a set of *administration views* of the *organisation model*. Each *view type* is implemented by a *generic* HTML page on the WWW server. A *view* (*an instance of a view type*) is implemented by a *specific* HTML page (*an instance of a generic HTML page*) displayed on the client side. A *CGI script* works as a *factory* [] by instancing generic pages in specific pages containing the required data extracted from the Co-Learn objects base. This *extraction* is performed by a *method call* (*to a Co-Learn object*) which has been created from a request translated by a dynamic interface invocation mechanism. The *parameters* of this request are the field values of a *form* (*included in a specific page*) filled in by a user. Thanks to an *authentication mechanism*, the Co-Learn object, which will perform the method call, knows the identity of the requester and can adapt his *reaction* by merging the requester *roles* and the organisation *rules*. [Fig. 1] below shows the links between the components and mechanisms described above.

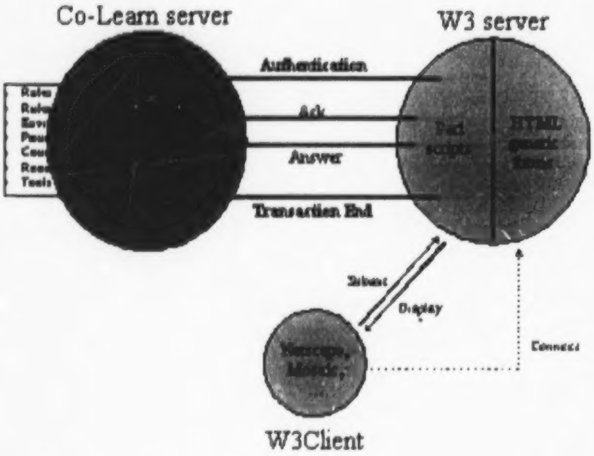


Figure 1 : The Co-Learn WWW gateway communication architecture

IV. Conclusion & Perspectives :

The Co-Learn users can now use the Web to configure their co-operation environments. They can also use it as an animation assistant during their co-operation sessions by dynamically setting up the tools, participants and roles involved in the virtual room. This functionality is offered thanks to a Dynamic View Generation mechanism which allows to inspect and to edit the data encapsulated in Smalltalk objects representing the organisation.

A drawback of the actual implementation using HTML forms and CGI scripts is at the interface level. Since we decided not to use frames because they are not included in the last HTML version (*and then not supported by all WWW browsers*), it was impossible to enhance the interface with Smalltalk-like browsers (*which would better merge global and detailed views*) in the management pages. Using the Java language should avoid this difficulty and allow to maintain durable connections between the client and the Co-Learn server. This way we could support group feedback.

The system we described here has been implemented to be integrated with a courseware development and publishing system based on HyperG [D'Halluin et al. 1996] to constitute a multimedia distance education global environment on the web, in the framework of the MODEM European project. The Co-Learn side of this global environment will manage the co-operative activities space and the HyperG side will manage the documents space. During the next months, we will enhance this Co-Learn browser to better support the time dimension of the activities space and to allow the management of technical resources (e.g. audio-bridges). In the long term, this tool will allow meta-management by supporting design, development and integration of new co-operative activity types and roles in the Co-Learn object-oriented framework.

V. References

[Madsen 1989] C. M. Madsen (1989) Approaching Group Communication by Means of an Office Building Metaphor. First European Conference on Computer Supported Cooperative Work (ECSCW'89). 449 - 460.

[Hammainen, Condon 1991] Heikki Hämmäinen, Chris Condon (1991). Form and room : Metaphors for Groupware. COSC 91/ACM. 95-104.

[Benford 1991] S. Benford ***(1991).*** A Real World Metaphor for a CSCW Environment. CoTech WG2 Internal Document.

BEST COPY AVAILABLE

[Engelbart 1992] Douglas C. Engelbart (1992). Toward high-performance organizations : a stratgeic role for groupware. CSCW 92 Proceedings. 77-100.

[Hoogstoel 1992] Frédéric Hoogstoel (1992). Au croisement de trois interfaces ... l'assistant d'animation : un agent informatique au coeur d'un collecticiel, in [TELEC 1992].

[TELEC 1992] TELECOM Paris (Eds) (1992). Document de travail des quatrièmes journées sur l'ingénierie des Interfaces Homme-Machine IHM'92. TELECOM Paris 92 S 004.

[Fitzpatrick, Kaplan, Tolone 1995] Geraldine Fitzpatrick, Simon M. Kaplan and William J. Tolone (1995). Work, Locales and Distributed Social wOrlds. 1995 European Conference on CSCW (ECSCW '95), Stockholm. 1-16.

[Hoogstoel 1995] Frédéric Hoogstoel (1995). Une approche organisationnelle du travail coopératif assisté par ordinateur. Application au projet Co-Learn. Data Computer Ph. D. Thesis, Université des Sciences et Technologies de Lille, France.

[Appelt 1996] Wolfgang Appelt (1996). CoopWWW - Interoperable Tools for Cooperation Support using the World-Wide Web. ERCIM Workshop on CSCW and the Web, Sankt Augustin, Germany.

[Dix 1996] Allan Dix (1996). Challenges and Perspectives for Cooperative Work on the Web. ERCIM Workshop on CSCW and the Web, Sankt Augustin, Germany.

[Whalter 1996] Michael Whalter (1996). Supporting Development of Synchronous Collaboration Tools on the Web with GroCo. ERCIM Workshop on CSCW and the Web, Sankt Augustin, Germany.

[D'Halluin et al. 1996] Chantal D'Halluin, Stéphane Réthoré, Bruno Vanhille, Claude Viéville (1996). Designing a Course on the Web : The Point of View of a Training Institute. WebNet'96, to be published.

Avoiding Vendor Lock-in: An Open Approach To Building Internet-Based Applications

Rick Horowitz
Internet & EC Consulting Practice
SPL WorldGroup Consulting
U.S.A
rick_horowitz@splwg.com

Introduction

For the first time in the history of the commercial computing industry, users of computer systems have the "real" option to build commercial-grade business applications using open technologies available from multiple competing vendors. The savings from such a strategy should be great, as multiple vendors will be forced to compete for the hardware platform, system services, communications infrastructure, and development tools used to enable the next generation of business applications. This short paper discusses the obstacles that vendors are placing in our way to prevent us from true vendor independence.

Last Year's Battleground

Last year's standards battleground revolved mostly around Netscape, as they "improved" on the standards with their own proprietary features. Netscape both won -- and lost -- this battle. They won by influencing the direction of the web, setting de facto standards for protocols and features, and increasing the overall usage of the Internet. They lost because they have been largely unable to lock customers in their proprietary extensions. Their key competitors, including Microsoft, have implemented the most important Netscape extensions, including many of Netscape's proprietary APIs.

1. HTML: the document description language for hyperlinked web documents. Netscape created many non-standard HTML extensions, restricting HTML pages designed with such features to work well only with Netscape Navigator. However, now that Microsoft has implemented virtually all of the proprietary Netscape extensions, and has publically committed to supporting the HTML standards process, it appears that all browser vendors, including Netscape, will fall in line with the standards. The tag line, "Best viewed with Netscape Navigator", is not seen nearly as frequently now as in the past. Stick to the HTML standards in your apps. HTML will not be a future battleground for vendor lock-in.
2. HTTP: the protocol used between a web browser and a web server, allowing a browser to request downloads of web pages, and enabling the transmission of HTML form data to web-based applications. This battleground is quiet and should not become a source of vendor lock-in.
3. S-HTTP and SSL: two separate "standards" have emerged for secure (i.e. encrypted) transmission between the web browser and web server. S-HTTP, promoted by the CommerceNet consortium, has received fairly wide support, though SSL, from Netscape, has received even broader support, and should be considered a *de facto* standard.
4. CGI, NSAPI (now called Netscape Server Plug-in API), and ISAPI: CGI is the "standard" means of interfacing an application with a web server, allowing the application to receive and process user data from HTML forms, and allowing the application to transmit dynamically created HTML pages back to the user's web browser. Although CGI is universally supported, its performance is poor. For this and other reasons (e.g. vendor lock-in and functionality enhancements) Netscape and Microsoft have created alternative mechanisms for interfacing a web-based application with a web server -- NSAPI and ISAPI, respectively. So far, although these interfaces have been implemented by some other web server vendors, neither Microsoft nor Netscape have implemented the other's server API. Use of NSAPI or ISAPI could restrict future application portability between web server vendors -- Caution!!

5. Netscape Client Plug-in API: Netscape invented the plug-in API in 1995, allowing downloaded *native* applications to operate within the browser context, providing a greater level of integration between the browser and the application than integration as *helper* applications could. The plug-in API has been adopted and supported by Microsoft and some other browsers, making it usable as a *de facto* standard.

The Next Battleground - Java

The new battleground is Java. "Hey, wait a minute", you might ask. "Sun created Java -- don't *they* control it?" Well, yes -- and no. Sun controls the name and they control the definition of the core Java libraries. But they don't have any control over additional class libraries that other companies might happen to make available to developers of Java applications and applets.

Contrary to many people's expectations, Microsoft is now marketing Java heavily, urging their customers to use it. What, then, is the problem? The (potential) problem is that Microsoft is marketing Java as nothing more than a language with which to build Microsoft-proprietary ActiveX controls (formerly OCXs). The advantages, say Microsoft, are cross-platform execution, due to Java's platform independence, *and* automatic integration with Microsoft's environment. Microsoft is even taking steps to create an *independent* standards organization to make future ActiveX direction open to the industry, and claims that we will soon see distributed COM (the object communication underpinning of OLE/ActiveX) on multiple platforms, including Unix and Mac. *Extreme* caution is advised, however. Microsoft has a history of creating and controlling standards for its own benefit, generally to the detriment of its competitors and even customers. This looks, smells, and sounds like a lock-in strategy. Beware!

What about Netscape? Well, we like Netscape's strategy a lot better than Microsoft's, but don't believe that it is benign with regard to lock-in. Netscape's standard is based on a conglomeration of technologies and APIs recently renamed Netscape ONE. Some of the technologies found in Netscape ONE, such as the client plug-in API, are supported by multiple vendors. Importantly, Netscape's system services, including security (based on X.509), messaging/email, and directory services (based on LDAP), are absolutely open and will be supported by many vendors. Our main concern with Netscape's strategy is the core of their application development strategy, a new and to-be evolved set of Java class libraries, dubbed the Internet Foundation Classes (IFC). IFC will ultimately provide access to the full range of Netscape services for Java-developed applications via the IFC APIs. Many of the planned libraries appear to overlap with offerings in the works from Sun, including the initial offering within the IFC umbrella, a UI library, that directly overlaps with AWT, the core UI library from Sun. Although Netscape is licensing the technology to other vendors, which should, in theory, even the playing field, Netscape will hold a performance advantage over other vendors unless licensees build IFC into their browsers, and could develop an unbreakable competitive advantage over the rest of the marketplace by controlling the interfaces through which applications derive system services -- ala Microsoft, today.

What about Sun Microsystems? Sun's JavaSoft unit has been very busy defining a broad set of class libraries (i.e. APIs) for important services including database access, security, payments, and the all-important JavaBeans effort to unify all the major object document standards, including OLE/ActiveX, OpenDoc, and LiveConnect (Netscape). Since Sun is freely licensing its Java technology for inclusion in other companies' products, including both Netscape and Microsoft, use of Sun's base classes appears to be the safest and most open choice for the future. All licensees of Java are required to support the entire core library, ensuring technical support, if not marketing support for these evolving Java standards.

Conclusion

So, what is one to do? If one is happy to be a captive of either Microsoft or Netscape, the answer is easy. Put your hands behind your back and enjoy your new handcuffs. They shouldn't hurt any more than the old ones. On the other hand, if one prefers an open approach that will not give you the vendor lock-in blues, investigate which vendors are supporting which standards, align with vendors and consultants dedicated to truly open applications, maintain a bias towards using Sun's Java APIs, start some pilots, educate, educate, educate, and keep smiling. For in the long run, you will come out the winner.

Web Courses for Northern B.C.: Virtual Community, Virtual Pagination

Ian Howatt, English Programme, The University of Northern B.C.
ian@andreae.unbc.edu

Basia Siedlecki, English Programme, The University of Northern B.C.
basia@donne.fac.unbc.edu

Project Objectives

In conjunction with the Open University Planning Council, and the Innovations fund of the province of B.C., we are part of a team responsible for the development of four on-line English courses that will be offered through the University of Northern British Columbia (UNBC). In recognition of the promising pedagogical applications of the World Wide Web, the principal objective of the project was to effectively replicate the classroom reality while simultaneously incorporating the Web's many instructive advantages.

To achieve this, we began with the construction of a virtual campus (in image-map format) that allows the student on-line access to all facets of the education process a student physically present at UNBC would encounter. In this respect, then, he or she can talk with the registrar, sign out books from the library, meet with a writing tutor, and so on. Perhaps most importantly, in order to facilitate the growth of a virtual community, we have constructed a number of communal "chat" areas for users, both at this introductory level and within the classroom situation itself. In each, students are provided with a means by which to talk amongst themselves with or without the professor's virtual presence. The sense of community offered by such a mechanism, we have learned, is vital to not only the learning process but also to the success of the entire project in terms of student failure and/or drop-out rates.

From this introductory interface (one we feel best accomplishes our primary goal and maintains a large degree of user-friendliness), the student (virtually) enters his/her chosen class(es). Although each has been designed by a different student under the supervision of an instructor, each contains a number of similarities that, we feel, distinguish this project from other similar ventures.

Before turning the podium over to Basia Siedlecki who will explicate these unique features in further detail, I would like to take a few more moments to describe the front-end of my particular work (simply because I am most familiar with it) -- English 420/620 The Literature of First Nations. In order to best employ the pedagogical advantages offered by the Web and in following the parameters set before us in the initial stages of the project, we employed a number of distinctive elements. For instance, each page presented to the student contains directional links that provide the user with quick and easy access to all components of the class, including an on-line MLA guide, a reference section on plagiarism, a large bibliography that includes on-line resources, the grading system, and so forth. All essay topics are on-line and the student's final paper is to be submitted by E-mail in conjunction with the shipment of the work in hard copy by snail mail. While this latter decision somewhat overthrows the initial objective of maintaining the virtual classroom throughout, a mailed hard copy is required to allow for complete and fair grading by the instructor. Student evaluation (i.e. grading) is conducted over the Web as is an introductory student information sheet and a final course evaluation form. Throughout each lecture (one per week, 13 in total), the student must respond to a series of questions designed to stimulate on-line class discussion (this aspect represents 30% of their final grade to ensure student participation). These discussion groups are threaded by topic and/or subject and, for research purposes, can be accessed at any time via a local search engine created by Dr. Stan Beeler at UNBC. Working in tandem with virtual pagination and the pre- and post-processing packages also designed at UNBC, this search engine offers an excellent means by which to provide an exciting and innovative educational opportunity to not only the North (with its limited Net resources) but also to those with the most sophisticated equipment anywhere in the world.

Technical Outline

The realities of developing on-line courses for a Northern University are far removed from the seemingly limitless possibilities of Java or Shockwave. But in tailoring our courses and our courseware development tools to the northern cyber climes, we have managed to develop several innovative solutions to our technical stumbling-blocks. Slow modems, bad phone lines, and old computers (without readily available support) necessitated a minimalist approach. We pared down the graphics to 16 colours, used clickable thumbnails throughout, made our imagemaps redundant line drawings and avoided audio and video altogether. Since courses inevitably seem to require lectures, we developed a PERL package that combines search engine and indexing functions with virtual pagination. The end product is paginated, for ease of reference for both students and instructors. As well, it provides a standard for on-line bibliographical documentation and reference, forcing page 3 to always be page 3, regardless of the screen size, browser type or font preference of the user.

The search engine, developed by Dr. Stan Beeler, (accessed on-line through an HTML form) addresses the slow modem problem by loading only the page requested. It uses a CGI interface to either load an index of the document in question or do a word Boolean AND search. Either way, the engine yields a list of "hits." Each hit is a hot link on an HTML page. Accessing any link will yield that virtual page of text, with the search term listed on top and bolded throughout. From there, it is possible to move forward or backwards in the database (like leafing through a book) or to go back to the home page (the search engine page). So, in response to any one request, only a single virtual page of data is being transmitted.

Part of the mandate of the project was to develop courseware tools to enable professors, who were novices at Unix computing, to produce effective on-line courses. Naturally, we felt that the virtual pagination search engine was something that was ideally suited to our cyber-milieu (the low tech north), so we created a preprocessing package to transform text into a format that the search engine could use. The preprocessor (this was my contribution to the project Basia Siedlecki) requires ASCII text, uses no command line arguments, sticking to a question and answer format instead, and allows for a very minimal knowledge of the Unix environment on the part of the user. It allows the course developer to specify the length of the virtual page and provide a custom header and footer if desired. It also provides for users with a knowledge of HTML, allowing them to insert custom tags, links, images, etc. An appended program transforms special characters from hex to HTML numerical tags if needed.

Once the text is preprocessed, it is ready for the search engine and needs only be loaded into the appropriate directory where it resides as ASCII text. Acknowledging the limitations of human endeavour, we have also provided for on-line post-processing or editing. This part of the project (developed by Dr. Beeler and embellished by Lynda Williams) is intended for touch-up work, adding images, correcting spelling errors, adding cosmetic HTML, etc. Essentially, it splits the text into paragraphs, putting each paragraph into an HTML form TEXTAREA, and allows for on-line editing. The resulting text is saved on the dedicated server under a different name, and can be viewed immediately following editing as a web page. Editing can continue until the course developer is satisfied with the resultant page. The new file is then saved under the old name and the old file is put into a backup directory. All this is done on-line, through a browser, without any back-end Unix work.

Our project was tailored for a very specific audience. The courses are aimed at users (students) with very limited computer resources and poor Internet connections. This, unfortunately, is a reality of Northern living. The courseware development tools are aimed at a more general spectrum of people, ranging from web novices to HTML experts. Internet distance education shows every indication of being the most practical and cheap solution to the problems of providing quality educational options to people in remote and isolated places. Our work is already being used as a paradigm for the development of distance education courses in the north. We are publishing our software package for general academic use and in the spirit of academic freedom will not be charging for copies.

Visit our site at http://donne.fac.unbc.edu

Acknowledgments:

Special thanks to the agencies that funded our project: The Open University Planning Council, The Innovations Fund of British Columbia and the University of Northern British Columbia. As well, we would like to thank Dennis Macknac and the Regional Operations Department at UNBC for providing support, encouragement and lab space. Finally, we would like to acknowledge the technical support and general guidance of Dr. Stan Beeler and everyone else involved directly and indirectly in the project

Harvesting the Internet: Creating a Resource on the Treatment of Skin Disease Using the Deliberations of an E-mail Discussion Group, RxDerm-L.

Arthur C Huntley MD
Department of Dermatology
University of California Davis
USA
achuntley@ucdavis.edu

Introduction

Using Internet hypermedia tools for education offers new and unique opportunities and challenges for the educator. To employ the advantageous aspects of the Internet involves developing and using new methods of information gathering and presentation.. In both medical practice and medical education, the Internet provides new opportunities for improvement in the access and transfer of information.

To understand and use the inherent advantages of networked computers, one can begin by their enumeration. The potential advantages begin at information gathering and extend through presentation. They include: immediacy, ease of collaboration, global sharing of resources, extended archiving of material, improved access, interactive presentation, feedback, ability to offer more than one path through material, ease of revision, ability to make extensive use of images, potentially low cost to viewer. Any improvement over paper resource will probably arise from the use of these advantages.

Thus report involves the creation and subsequent harvesting of a dermatology e-mail group developed to provide a source of discussion about therapy of skin disease. The result of this effort is a searchable on-line archive of treatment discussion for skin disease. In order to determine the value of this resource, the information generated was compared with that available from using Medline and from using DermRx (a software program developed and marketed by the American Academy of Dermatology).

Methods:

RxDerm-L was created as an e-mail list at the University of California Davis in November 1993. With the list description noting that discussion was intended for and by dermatologists. It was initially open for subscription, but subsequent to a few intrusive requests, subscriptions are moderated. Posting is only open to subscribers and comments are not moderated.

Beginning in June 1995, archiving of the list discussions was initiated, and the generated material filed by disease. The letters are not edited except for occasional off-topic comments. The archives were first made available by gopher, and subsequently on a WWW server (http://matrix.ucdavis.edu).

For the purposes of evaluation, a list of 50 diseases was drawn at random from the index of a common dermatology textbook. The output for each disease was then compared using RxDerm Archives, DermRx, and Melvyl Medline. Comparisons were made about the number of treatment options elicited by each program for each disease and the depth of the discussion. An attempt was made to estimate the quality of information and any obvious deficits.

Results:

At the time of this writing, the RxDerm-L list has more than 300 members in 30 countries. The members are almost all dermatologists. From June 28, 1995 until June 8, 1996, there were more than 1800 letters, most of which were on the stated topic of the list, therapy of skin disease. The number of postings has gradually increased and currently often exceeds 20 per day.

The number of diseases archived at the time of this writing is 207. The discussion on various diseases can range from a few paragraphs to several pages.

At the time of the initial survey, approximately half the diseases on the list were not yet part of the RxDerm Archives. This compares with 10 that were not part of the DermRx program and that were not found under that name in Index Medicus. For those diseases found on all three resources, DermRx had the most number of treatment options, RxDerm Archives had the most discussion, and Melvyl Medline the most up-to-date information.

Discussion:

The Internet is spawning new forms of dialog. Persons with similar interests worldwide are now connected and discussing information of mutual interest. In the case of dermatologists, the e-mail list termed RxDerm-L is used for practitioners to discuss alternative treatment options. Judged by the daily traffic on this list, it is successful.

This method of information gathering employed here is different from traditional library research. This project uses several of the enumerated advantages of networked computers. Immediacy, perhaps the greatest advantage of e-mail, is the key to facilitating this type of dialog. The accumulated practice knowledge of hundreds of networked specialists is focused as a collaborative effort on discussing the treatment of specific diseases. The discussion is archived to create a disease-oriented and searchable text for clinicians.. This resource is intended to serve as a freely accessible reference for practitioners. Access is easy with on-line search and retrieval. The process of revision occurs continuously within the context of an on-line discussion group.

Information harvested by this method tends to be much more anecdotal than that found in Journals and Texts. It is subject to the biases of a multitude of authors writing under minimal editorial restraint. As such it is best interpreted by the physicians who are already trained in this specialty. Nevertheless, it appears to provide a useful resource for practicing dermatologists. For many diseases there is an in depth discussion about patient management. For most letters, the authorís name and e-mail address are present, so the reader might access the author for further clarification..

The future course of the RxDerm Archives probably includes the addition of editorial oversight and some additional search capabilities. As the discussions become lengthier, the moderator will need to select responses for inclusion into the Archives. Editorial pruning will delete duplicated remarks, and limit unsubstantiated comments. The added search capabilities will allow the reader to search by drug across disease categories.

This project has demonstrated that a popular therapy e-mail list can be harvested to produce a useful reference resource for the involved community. The potential of harnessing a portion of the global expertise of a medical specialty should prove a powerful incentive to further projects of this kind.

School Participation in the Co-Development of Commercial, Web-based Curriculum Resources

Penni Leigh Jones
Allen Independent School District
Allen ,Texas USA
0007102301@mcimail.com

As one of the fastest growing school district in the state of Texas, Allen ISD is committed to providing the students in our district the strong academic foundation they need for success. The challenge in maintaining these quality standards is the ability to adapt to change and the need to guide this change in directions that will support the diversity of learning styles and maximize learning and productivity. Allen ISD is determined to reengineer the educational experience by creating a community of learning with a strong focus on technology.

In response to changing needs, technology is distributed in a variety of ways from laptops for each of our academic staff, pods for open access, lab settings for curriculum integration, and literacy labs to teach skills. Our students and staff can access a variety of productivity, research, and telecommunication tools through our local and wide area networks. Our T1 connection to the Internet allows student and staff to be publishers and use the rich resources found throughout the world. Through EMG, our students go on electronic field trips and utilize customized curriculum. Our Josten's labs assist in TAAS preparation. In addition, our high school students will be working with a software company this year to produce high-end multi-media products. Our SHIFT (Students Helping Integrate Future Technology) program at our schools produce web pages for staff and train staff and students on how to use powerful technology applications, such as Microsoft Office, Adobe Photoshop, Adobe Premiere, and Director.

We have developed a K-12 technology curriculum to provide students opportunities to master higher level skills, to practice effective communication, to work individually and collaboratively on real-world tasks. Our instructional designers on each campus provide staff development to change existing teacher roles to mentors that would facilitate direction and motivation of students and fully implement our technology learning goals. We also hosted a symposium " Technology Leaders and Users" to develop administrators into more effective and efficient managers of instructional resources, communications, and school-community relations, while expanding methods of assessment to assist diagnosis, provide feedback for instruction and support accountability.

As leaders in instructional technology, we have started many collaborative projects that enrich our students' experience. We have helped telecommunications companies understand its impact in education. We have assisted product development companies produce solutions for educational markets. In addition, our students are creating software with companies and are training local business on software applications. These collaborative efforts are making a difference for our kids, who will in turn impact their twenty-first century world.

Current Situation of Educational Utilizations of Computer Networks in Korea

Sook-Hi Kang
Center for Educational Multimedia R & D
Korean Educational Development Institute

In the present "Information Age" which requires a fundamental change in educational environments, there has been a growing need for the use of computer networks to enhance the quality and effectiveness of teaching and learning in school and the efficiency of school administration. In considering that most countries in the world today seek to promote their international competitiveness through education, the utilization of the current information and communication technology for educational purposes now emerges as the most pressing task in Korea.

In this context, this study was initiated as part of the Korean-led APEC educational project which focuses on the utilization of computer networks for education. More specifically, this study was conducted as a pilot study of the APEC project which aimed at investigating the current status of the use of computer networks in teaching and learning, school administration, teacher training, etc. in Korea. The main purpose of this study was to establish basic information and data concerning the current status of the utilization of computer networks in schools in Korea. For this purpose, three different types of questionnaires were developed, each for teachers, students, and computer teachers.

The areas of survey for teachers included computer network experience, teacher training, efficiency of communication through computer networks, future plan for using computer networks for instructional purposes, and obstacles to the utilization of computer networks. For students, the survey was about computer network experience, utilization of computer networks for learning, and future plan. The areas of survey for computer teachers included the situation of computer network equipment/facilities in school, teacher training, utilization of computer networks in school, effectiveness of the use of computer networks, outside projects/policies for computer network utilization, side-effects of the use of computer networks, management, obstacles to the use of computer networks, and outside support.

For this study, twenty-two schools using computer networks were selected on a random basis. Among them, there were 14 elementary schools, 2 junior high schools and 6 high schools. For each school, it was encouraged that one teacher for each different subject, 20 students per grade level, and one computer teacher would participate in the survey. The surveys were conducted and the results were analyzed. For each school, a face-to-face interview with the computer teacher was also held.

The results of the survey indicated that currently the utilization of computer networks in Korean schools is still at a very infant stage. Not many teachers have used computer networks, especially for instructional purposes. Although it was found that there were many factors contributing to the low utilization of networks in school, several of them can be pointed out. For one thing, schools are not sufficiently equipped with needed hardware and software. For another, there are very few examples of the good use of computer mediated communication which teachers can follow. That is, teachers do not have knowledge about how to apply the available technology to teaching-learning activities. Some other barriers to teachers' low use of networks included not enough institutional, organizational, and curricular support, low level of computer network connection, lack of modems and exclusive communication lines, lack of teachers' technical ability to use computer networks, and so forth. However, most teachers expressed that they have a positive attitude about the computer mediated communication and have a plan to use it more often especially for instructional purposes.

The survey also showed that students use computer networks more at home than in school. One probable cause seemed that because computers are mostly located in the lab not in classrooms, they were not easily accessible especially for students. Some other barriers to the students' low use included lack of facilities and equipment and lack of time to use computer because of their busy study schedule. However, students expressed that computer networks can be helpful and effective for their study and may build their information technology capability and close relationship with teachers.

It is hoped that the results of this study be used as a resource material for promoting the use of computer networks in schools by providing school teachers and administrators with various information on how to utilize the networks for instruction and administration. It is also expected that this project will provide a basic framework for adopting technology into schools and other educational environments, seeking educational improvement using technology, and sharing educational information with other countries.

Acknowledgements

Thanks to our Korean-led APEC team members for letting me have this opportunity to present this paper.

The Web as an Instrument in Redesigning the Teaching/Learning Transaction

Cathy C. Kaufman, Ph.D.
Professional Studies in Education
Indiana University of Pennsylvania

The World Wide Web is a pivotal component in educational transformation, one that affords teachers (K-12) the opportunity to direct significant curricular and pedagogical changes. Using the World Wide Web in teacher preparation courses motivates pre-service educators to use a wider range of pedagogical approaches in their instructional planning and to support student learning with access to a wider variety of materials.

Traditional approaches to learning have been grounded in three basic orientations: 1) learning formal disciplines, 2) learning general skills, and 3) learning domain-specific knowledge. More recently, however, Perkins and Salomon [Perkins and Salomon 1989] suggest that how we structure the thinking and processing conditions surrounding learning opportunities is equally as important as content knowledge. The structure of the World Wide Web parallels this new thinking in cognitive science in a way that few instructional resources can claim. Through its linking and cross-referencing capabilities, this technology naturally fosters higher level thinking skills such as synthesis of information. And it gives teachers the resources to encourage self direction, personal motivation, and application of formal schooling to real world context. Students and teachers alike can expand their own knowledge base and demonstrate the value and the necessity of life long learning. Technologically integrated learning opportunities in classrooms encourage cooperative efforts, individual responsibility, and team learning.

Teacher education is a central element in charting new directions in teaching and learning. Without the capability to access knowledge in new ways, to critically process that information, and to establish motivating and rewarding learning experiences, novice educators are vulnerable to seeing little or no connection between their academic training and their changing roles and responsibilities in twenty first century classrooms. The World Wide Web is one possible instrument in preparing novice educators:

- who will develop technological approaches that enable them to keep pace with the constant evolution of content information,
- who can model critical thinking skills and problem solving abilities into their own classrooms, and
- who will transform rather than conform to present practice.

References

[Perkins, D. and Salomon, G, 1989]. Perkins, D. and Salomon, G. (1989). Are cognitive skills context-bound? Educational Researcher , 18, 16-25.

COMMUNITY WEB PROJECTS TO SUPPORT HUMAN-COMPUTER INTERACTION EDUCATION

Carol Kilpatrick
Department of Mathematics and Computer Science
Georgia State University
Atlanta, Georgia USA
matcek@cs.gsu.edu

Overview and Context

Rationale, goals and experiences related to class Web projects with community organizations are described. The projects were done over the past eight months to support learning in Human-Computer Interaction (HCI) classes at Georgia State University (GSU). GSU is a large, urban university located in downtown Atlanta. HCI class size is usually 15-20 students (most students are computer science majors). Projects usually involved teams of two students and two members from an Atlanta area organization.

Why Community Web Projects in HCI Education?

There are a number of reasons the Web is useful in HCI education. Interaction possibilities on the Web are increasing. Students can apply results and techniques from HCI research and practice in their Web projects. They can modify and reflect on these results and techniques within the context of the new technology of the Web. Current professional involvement within the HCI community provides an excellent model for student involvement with usability and the Web. Examples include workshops on HCI and the Web at CHI96 and Hypertext96, the recent symposium, The Missing Link: Hypermedia Usability Research and The Web and the monthly Alert Box column by Jakob Nielson of Sun Microsystems.

Project Goals and Examples

Six fundamental goals provide the framework I used for designing the structure of these HCI projects. Each goal is listed followed by examples of how these goals were met with the community Web projects.

1. Provide student exposure to actual users. The HCI student role is to bring technical expertise about interaction and usability to a project. The community participant role is to bring user, domain expertise. The community participants provided concrete, relevant experience for students to test various HCI techniques associated with user-centered design. In addition, the projects provided a springboard for consideration of social and organizational issues (important considerations in HCI) related to the Web.

2. Provide students with experience concerning current HCI issues and methods. A goal of class projects is to provide concrete experience that supports student understanding of course content areas. The following examples describe HCI content areas incorporated.

One team used a technique discussed in the design of Sun's Web site. The technique is used to generate a user model of an information space and is based on a card sorting technique outlined in the Sun Web site discussion. Several class teams used a thinking-aloud protocol [Nielson 94] for user observation. For example, one team worked with a local library to design a prototype document search application. They used a thinking-aloud protocol to observe use of several similar Web-based search applications currently in use at other libraries. One class did an exercise using heuristic evaluation [Nielson 94] and then read an article relating to heuristic evaluation of Web sites, "Guidelines for Designing Usable World Wide Web Pages" [Borges et al. 96]. A team then led student heuristic evaluations of several student Web applications.

Winter and summer quarters, all class teams used a software design technique based on the participatory design [CACM 93] method, PICTIVE [Muller 93, Muller et al. 95]. With PICTIVE, design team members use common office supplies such as paper, index cards, felt-tip pens, sticky notes, etc. for group design work (see several PICTIVE-based student examples). PICTIVE promotes mutual learning and equity in expression by all members of a design team. Students and community participants found PICTIVE an effective communication vehicle during their Web project design.

3. Incorporate service learning. "Service learning connects young people to their community, placing them in challenging situations where they associate with adults and accumulate experiences that can strengthen traditional academic studies" [Service Learning 93]. Students chose projects from a variety of community organizations including, a local geriatric center, a group involved with medically fragile children, the Georgia Lung Association, a small elementary school integrating the arts with academics and a group providing technology for the disabled. Much of the initial inspiration for incorporating a service component in HCI course work came from work by Jean Gasen [Gasen 96] at Virginia Commonwealth University. Her efforts involve incorporating social responsibility in a collaborative course involving a local cancer center and university faculty from the School of Art and the School of Business.

4. Emphasize and practice communication skills. When working with student teams, it is a mistake to assume teamwork skills do not have to be taught. Throughout project work, attention was given to skills such as conciseness, listening, reflection and equity in participation. A primary motivation for this community project work is found in the 1994 Report on New Directions in HCI Education, Research and Practice [Strong et al. 94]. The reports calls for an increase in student exposure to problem situations that challenge them to "learn to work with others, and to justify their choices to professional colleagues.

4. Provide students multidisciplinary exposure. Many issues in HCI are multidisciplinary by nature. Multidisciplinary teams are required in the development of many software projects. The community projects provide students HCI experience working with users and experts outside their discipline.

Conclusions

Karat and Dayton raise the issue "...usability activities must be practiced by more members of the design and development team than just the usability specialists" [Karat and Dayton 95]. The community Web projects reported here address this issue through education involving hands-on experience with usability activities. As possibilities for interaction increase, there is a need for more HCI usability practice on the Web. Community Web projects promote student learning about HCI in general and also about HCI and the Web. These projects also support user involvement in learning more about HCI and the Web.

References

[Borges et al. 96] Borges, J.A., Morales, I. and Rodriquez, N.J. (1996). Guidelines for Designing Usable World Wide Web Pages. CHI'96 Conference Companion: ACM Conference on Human Factors in Computing Systems, Vancouver, Canada. 277-278.
[CACM 93] Special Issue on Participatory Design. (1993). Communications of the ACM, 36(4).
[Gasen 96] Gasen, J. (1996). Encouraging social responsibility through collaborative team learning. CHI'96 Conference Companion: ACM Conference on Human Factors in Computing Systems, Vancouver, Canada. 125-126.
[Karat and Dayton 95] Karat, J. and Dayton, T. (1995). Practical education for improving software usability. CHI'95 Conference Proceedings: ACM Conference on Human Factors in Computing Systems, Denver CO. 162-169.
[Muller 93] Muller, M. "PICTIVE: Democratizing the Dynamics of the Design Session". In *Participatory Design Principles and Practices*. 211-237. Douglas Schuler and Aki Namioka, Eds. (1993). Hillsdale, NJ: Lawrence Erlbaum Associates.
[Muller et al. 95] Muller, M. J., Tudor, L. G., Wildman, D. M., White, E. A., Dayton, T., Carr, B., Diekmann, B., and Erickson, E. D. "Bifocal tools for scenarios and representations in participatory activities with users". In *Scenario-Based Design for Human-Computer Interaction*. 135-163. J. M. Carroll, Ed. (1995). New York: Wiley.
[Nielson 94] Nielson, J. (1994). *Usability Engineering*. Boston, MA: AP Professionals.
[Service Learning 93] Standards Committee. (1993). The Alliance for Service-Learning in Educational Reform.
[Strong et al. 94] Strong, G. B., with J.B. Gasen, T. Hewett, D. Hix, J. Morris, M. J. Muller and D. G. Novick. (1994). A Report: New Directions in Human-Computer Interaction Education, Research and Practice. Sponsored by: NSF Interactive Systems Program, NSF Applications of Advanced Technology

Program and Advanced Research Projects Agency Software and Intelligent Systems Technology Office.

Automatically Generated Web-based Multimedia Presentations

Dr. Richard L. Klevans
Department of Computer Science, North Carolina State University
Box 8206, Raleigh, NC 27695, Email: rlklevan@eos.ncsu.edu, Telephone: 919-515-6535

Dr. Mladen A. Vouk
Department of Computer Science, North Carolina State University
Box 8206, Raleigh, NC 27695, Email: vouk@adm.csc.ncsu.edu, Telephone: 919-515-7886

Overview:

Using existing technology, such as Web browsers, HTTP servers, HTML documents, CGI programs, and RealAudio™, we have developed a system for automatically generating and serving low-bandwidth Web-based multimedia presentations based on live versions of the same presentations. This system has many potential uses such as delivering distance education through on-demand lectures, and capturing conference presentations and making them available to wider audiences.

To make these presentations available via the Internet to the widest possible audience, a low-bandwidth solution is necessary. A presentation, whether it be a lecture in an educational setting or a presentation at a conference, typically involves of a person who talks about a set of slides or transparencies. Most of the information is in those slides and in the presenter's speech. The image of the talking head contains little information relative to the content of the presentation. Hence, high-bandwidth video is often not required to successfully transfer the presentation information. We use HTML versions of the slides and the accompanying audio to reduce bandwidth but still retain a high percentage of the original presentation information. Synchronized audio with automatic page turning is used to simulate the original presentation as closely as possible. However, when the situation demands it, we also use high-bandwidth video clips. A user can watch our presentations in a hands-free fashion, but he/she can also ìstop, pause, rewind, and fast-forward ì the presentation in a ìrecorder/player-typeî fashion.

In order to minimize the effort required to transfer presentations onto the network, tools for automatically generating the Web-based presentation from the original presentations are necessary. One key piece of information is the timing data that describes when to advance to the next slide (or backtrack to a previous one). Our system provides mechanisms for automatically capturing the timing for the slides as the original presentation is being delivered, or in a post-processing mode if the system was not used during the original presentation.

Design Platform:

One of our goals is to make the presentation information available to as wide of an audience as possible. In order to do that, we found it necessary to use only existing and readily available technology. Standard Web browsers and HTML documents were a natural choice. To allow delivery of the presentation slides in multiple formats, such as with or without frames, each slide's content is kept separate from the control links that permit navigation through the lecture-e.g., the "Next" and "Previous" buttons.

For the audio portion of the presentation, we have chosen the RealAudio format for several reasons. RealAudio servers can transmit high-quality audio in a streamed fashion with minimal start up cost (Typically, 5 seconds or less). By "streamed" audio, we mean that the entire audio file need not be downloaded to the client's machine before the audio starts playing. Due to the high compression rate of the RealAudio format, the required disk storage space and network bandwidth are minimized. Even clients with 14.4 Kbps modem connections can successfully receive RealAudio files. RealAudio also provides a mechanism for triggering the client's browser to display a sequence of HTML pages based on time stamps. This gives the user of the system the illusion of being at the actual presentation as the slides automatically advance when the presenter moves on to the next topic. Furthermore, RealAudio players are available for a number of hardware platforms, and can be embedded into HTML pages as a Netscape plug-in..

Description of Current System:

Each presentation consists of a directory in the HTTP server path. The directory contains a presentation information file, a set of HTML files, and RealAudio information. The HTML files will contain the presentation. They are normally located in the same directory as the presentation information file, but this is not a requirement and they maybe located elsewhere on the some other server. The presentation information file allows the author to specify the order in which the slides are presented (specified as URLs) and information that appears on the presentation home page, such as presentation title, author, date, abstract, and links to other relevant information on the Web. An on-line editor allows the author to modify this information.

When the system builds a presentation, the *glue files* necessary to bind the audio-visual components together are created automatically. The glue files consist of HTML files that specify the frames and control links, the RealAudio metafiles (.ram and .rpm) for describing audio segments, and a RealAudio synchronization file (.rae) for performing automatic page turning. A printable version of the entire set of slides is also created. The *Preliminary Build* of a presentation creates the necessary glue files for two authoring modes.

Live Capture Mode is used when a presentation is given for the first time. The presenter uses the HTML pages prepared in advance as the slides for the presentation. The slides are displayed to the audience via large-screen TV set or an LCD projector. This equipment is becoming more commonly used in the conferences and is also becoming more affordable to educators. Naturally, the computer used for displaying the HTML pages must either be connected to the network or be a standalone HTTP server. It is assumed that the audio is captured by an auxiliary program and converted into RealAudio format either on the fly, or after the fact. The presenter starts the auxiliary recording program, and then begins the presentation. The presenter uses the "Start", "Next", "Previous", and "Done" buttons to navigate through the presentation. As this navigation occurs, a CGI program on the HTTP server records the times at which each slide is viewed.

If the constraints of the Live Capture Mode cannot be met, or if the presentation has already been given, the semi-automated *Post-processing Mode* can be used. For this mode, the content of the slides must be converted into HTML format. Many programs exist for converting other presentation formats into HTML, such as Microsoft PowerPoint and Word Internet Assistants and Latex2html. The audio files, which can be digitized either from audio or video tapes of the original presentation, must be converted into RealAudio format. With the HTML and RealAudio files in place, the Post-processing Mode allows a user, to navigate through the slides while the system plays the audio for the presentation. Each time a navigation button is pressed, the system will record the time in a data file for the presentation.

Once the timing information is collected and any corrections are made to the slides, a *Final Build* uses the timing information to create the necessary glue files for viewing the presentation.

Current Users:

The system is currently being used at the "Region Training Center for Parallel Processes" Web site (http://renoir.csc.ncsu.edu/RTCPP/) for delivering instructional material and conference presentations. It will also be used for teaching regular university classes by the Computer Science Department at the North Carolina State University starting in the Fall Semester of 1996. A demonstration of the system is at:

http://renoir.csc.ncsu.edu/RTCPP/HTML/Concepts/new/demo/index.html

Acknowledgments:

This project is supported, in part, by the NSF awards ACS-9418960, ACS-9418606, and ACS-9696131, and by the IBM Corp. SUR program. RealAudio™ is a trademark of Progressive Networks. PowerPoint

and Word Internet Assistant are products of Microsoft, Inc. Netscape Navigator™ is a trademark of Netscape Communications Corporation.

Computer Aided Information Navigation: Project Description

David Ribeiro Lamas*
Jennifer Jerrams-Smith
*Department of Computing, DEC
Bournemouth University, England
{dlamas, jj-smith}@bournemouth.ac.uk

Feliz Ribeiro Gouveia
*Multimedia Resource Center
Universidade Fernando Pessoa
Portugal
fribeiro@ufp.pt

Introduction

The Computer Aided Information Navigation (CAIN) project is proposed as a system that maximises information navigation quality on very large hypertext databases, such as the World Wide Web, without interfering with them. Maximising information navigation quality means to reach the desired information faster and more accurately trying to avoid common hypertext's user problems such as information and memory overload or disorientation. This will be accomplished by creating a local and collaborative environment which will act as an interface between users and the data infrastructure. This environment will be responsible by maintaining meta-information about the referred infrastructure as well as by maintaining information on users models and preferences and to use that same information to direct and optimise users' interaction with the system. As a way of describing the system without any ambiguity, Z was adopted as the formal notation for specifying the several system components. No operations will be described because World Wide Web creation and authoring are far away from this projects' aims and scope.

General architecture

At a general level the whole system architecture is decomposed into three layers. The data infrastructure - Level 0, in this prototype, will be the World Wide Web with its HTML pages and page embedded links. The information layer - Level 1 - is an automatically generated and tagged node and link database which represents the users known world. The basic purpose of this layer is to enable some sort of local hypertext manipulation and, eventually, visualisation of the known hypertext network. This automatically recorded database will also work as an enhanced navigation history keeper and will enable features as bi-directional navigation to and from any node on the hypertext network among others. The knowledge layer - Level 2 - is the core of the whole proposed architecture. At this level information about the nodes and links is stored enabling a generation of a sort of semantic network whose edges are information layer selected and frame described nodes and whose arcs are information layer selected and typed [Bloomfield 1994] links. This semantic network will, in some way, be combined with a user model [Beaumont 1994] database by a guiding agent [Brusilovsky 1995] now named the Butler. The Butler definition is the real aim of the proposed project.

Level 0's scope, however, is limited to the actual HTML pages and the related links. Within the pages themselves every link will be dealt with be it a link to another page, a link to a specific position on another page or a link to other parts of the current page. For a formal definition of the subset of the data infrastructure elements addressed by this prototype's architecture a Z definition will be provided. One set is assumed, [URL]. It represents the set of all possible Uniform Resource Locators all over the internet but for now only the HTTP addresses are relevant which leads to the definition of the subset:

Pointer: P URL

which contains only the selected range of addresses. It is possible to describe the World Wide Web as:

DataInfrastructure

node: Pointer © Content

title: Pointer © S

links: Pointer © P Pointer

-

dom node = dom title

dom links = dom node

where node relates each valid URL with its contents, title maps each valid URL with its title and links each node with a set of nodes.

The information layer is a shared node and link database created co-operatively by a group of users with similar interests. While they navigate the World Wide Web, the system records every step taken into a common local database. There are several problems in this related with consistency maintenance, co-operative browsing, ownership and privacy.

KnownPointer: P URL

-

KnownPointer z Pointer

MappedDataInfrastructure

knownNode: KnownPointer © Pointer

knownTitle: KnownPointer © S

knownLinks: KnownPointer © P Pointer

-

dom knownNode = dom knownTitle

dom knownLinks = dom knownNode

InformationLayer

knownNode: Pointer © Content

---

knownTitle: Pointer © S

knownLinks: Pointer © P Pointer

-

dom node = dom title

dom links = dom node

Final remarks

The next step on this project is to specify and implement Level 2. Besides the possibility of helping groups of users navigate in large information spaces through the sharing of individual experiences, CAIN also addresses issues of adaptive interaction. We hope to let the helper agents inside CAIN evolve so as to adapt to a user profile. Agents can follow each others user preferences when engaging in exchange of information. Information navigation is thus seen as a process of collaborative interaction.

References

[Beaumont 1995] Beaumont, I., Brusilovsky, P. (1995). Adaptive educational hypermedia: from ideas to real systems, H. Maurers, Eds., ED-MEDIA '95, Association for the advancement of computing education, Graz, Austria. 93-98.

[Bloomfield 1994] Bloomfield, H. (1994). Links in hypertext: an investigation into how they can provide information on inter-node relationships, PhD. Thesis, Queen Mary and Westfield Colledge, U. of London.

[Brusilovsky 1994] Brusilovsky, P. (1994). Adaptive hypermedia: the state of the art, P. Brusilovsky, Eds., MHVR '94, Moscow, Russia. 24-29.

Education for Changing Times - An Online Learning Framework

Iain Langdon, Senior Researcher
BT Labs, Martlesham Heath
Ipswich, Suffolk, UK, IP5 7RE
iain.langdon@bt-sys.bt.co.uk

Introduction

This paper describes research on Education and Training at BT Laboratories (BTL) in the UK. BT is the largest telecommunications company in the UK with 27.3 million exchange lines and had a turnover of £14,466m in 1995/96. It has a significant global presence - for example, a 20% share in MCI, the US long distance carrier. BT invests over £280m p.a. on research and development. Its Education and Training research programme addresses the many stages of the UK education system from primary schools through secondary and on to college/university. The training market is also being addressed by considering the needs of corporate users, small businesses and the increasing demand for services that support *life long learning*.

Programme Background

The approach is to work directly with the community to understand their needs and requirements. At the heart of the programme is a test-bed populated by real users whose usage can be observed and analysed. This *action research* approach produces results which feed into an iterative prototyping development cycle.

The aim is to explore the power and benefits of online learning, and to help BT understand how to manage and deliver these services. We are tackling the challenge of designing future technologies for Education and Training in several ways:-

- addressing the softer organisational issues as well as the more tangible technical problems;
- exploring how telematics (the convergence of telecommunication and computing) is applied in a number of different educational and vocational settings; and also
- addressing barriers to change, and developing a management of change strategy to overcome those barriers.

Common Components of the Programme

The Education and Training Research programme is given coherence through 3 major shared components:

The Learning Test-bed: This spans a wide range of computing platforms connected via access bandwidths from narrow through to broad band. While the core of the test-bed is situated within BT, significant parts of it exist in Suffolk County schools & colleges. The complexity and dispersed deployment of the test-bed provides us with credibility and validity, and ultimately leads to the design of more suitable and usable products and services. By using this test-bed, we will also develop an understanding of critical management of change issues.

A Common Learning Framework: It is widely acknowledged that there is a need for a new model of learning. This will involve more universal and flexible access to learning opportunities, allowing more effective and efficient pedagogy. In particular, it is becoming clear that dialogue is one of the keys to effective learning. Future educational systems must not only foster dialogue and be situated in meaningful tasks, but must also recognise that much of learning is a product of a social process, resulting from learners interacting with each other.

A Generic Learning Platform: From the Learning Framework it is clear that there are common elements to teaching and learning. Any technological solution should reflect these common elements by specifying generic building blocks which can be configured to meet the needs of specific education and training products and services.

The education and training systems of the near-future will exploit a telematics learning framework, utilising state of the art developments in video-conferencing, groupware, Internet, World Wide Web and distributed multimedia technologies.

The Project Exemplars

Our work involves a number of pilot projects within the community, some of which are described below.

The Televersity Concept

Close to BTL is University College Suffolk. The college aspires to become a 'Televersity' by the turn of the century. The vision is to provide a community based university which delivers pre-packaged material over the Internet, and electronic tutorial support using video-conferencing technology. As part of a feasibility study, we have set up some pilot studies. One study involves nurse education where problems of teaching dispersed and small groups is being studied. Another pilot, being run between the college and schools, considers School Centred Initial Teacher Training, where the focus is on using networked video-conferencing tutorial sessions.

Tomorrow's Customers

Tomorrow's Customers is aimed at schools (both primary and secondary). The overall aim of the proposed work is to 'visualise, show benefits for and evaluate with customers a range of medium and long-term educational solutions based on advanced video-conferencing services and the Internet'. Two of the main tasks are:

HomeLearn

The home is a key focal point for many education and training applications. The various interfaces that allow work to be carried out in the home are being explored, and a detailed scenario describing how education and training home services are used in a typical family is being developed.

CampusPlus

CampusPlus is exploring the key issues that need to be addressed by the next generation of BT's existing education sector product - CampusWorld. New technologies are being evaluated on the test-bed, using a story, "1000 years on the Orwell" (a river near to BTL) as a metaphor for communications. The example is chosen to maximise collaboration between schools, and to study them as they become content providers.

iCampus (Industry Campus)

This task explores network, service and training issues resulting from service delivery over high speed networks (Intranets). The training exemplars are initially being piloted within BT as a model of a typical corporate user. We use an Object Oriented (OO) approach, and are focusing on capturing and utilising personal preferences to produce more customised user interfaces. The OO findings are being fed into the other activities.

Acknowledgements

I am indebted to all of my colleagues in the BTL Education & Research Programme whose collective

ideas are summarised in this short paper.

Webbization In Russia

Dmitry Latukhin
Technical University of Chelyabinsk
Chelyabinsk, Russia
dl@urc.ac.ru

Networking space

Dynamic expansion of the Internet has affected boundless territory of Russia. Everywhere, from Moscow to the Far East, increasing interest is noticed in telecommunications and networking services and, in particular, in developing WWW sites. Sometimes this situation looks like mushrooms growing after a good summer rain. It depends on a lot of different factors including geography of the country, its information infrastructure, etc.

So far there is a moment restraining the development - this is extremely limited number of high-speed communication lines on the outside. Practically only Moscow and St.Petersburg have several links corresponding to T1/E1 ones at present. All the rest of cities are contented with slow lines - with no more than 19.2-28.8 kbps or 64-128 kbps via satellite channels in the best case. The explanation is in high monopolization of this sphere of business by few state and half-private companies. But because the local market of telecommunications is very attractive for large western companies and investors the situation is being improved step by step. New digital lines and earth satellite stations are built though they are too expensive for lease by as any state and educational institutions so originating firms-providers of Internet in Russia (for example, over $ 60,000 annual payment for 64 kbps INTELSAT channel).

Professionals

What is surprising that is high qualification and wide experience of Russian network staff. These people work under far worse conditions as compared to their foreign colleagues. Their sites have often PCs only and their salary is not more than $200-300. The system administrators and webmasters constantly face bad quality of leased lines and lack of any local firms supporting and developing software.

As a rule they examine a set of free and commercial operating systems (e.g. BSD clone, Linux, Solaris, etc.) and are able to modernize a HTTP daemon according to their requirements very quickly. These specialists have also additional «advantage» due to some time lag in technical level which allows them to turn towards choosing either approach and equipment. And as result a taken strategy and methods of implementation are usually correct and extensive. Total number of persons which are capable of handling access to the Internet and IP services is estimated as 20-30 people per ordinary Russian city except the capital one.

Dynamics and geography

In the end of 1994 your fingers were enough to account all Russian Web sites. Now, in the middle of 1996 there are almost 400. About 60-70% from them operate on regular basis. 70% of the servers are concentrated in Moscow and St.Petersburg where only 10% of the population live. Nevertheless you can find new interesting ideas also in Khabarovsk, Yaroslavl, and other large cities.

In the same 1994 few people or small and middle firms could venture to use direct IP or SLIP/PPP connections. The providers were rare and their services were too expensive (say, $200-800 monthly). Next year was a turning-point. Splash of activity was caused by realization of some projects on increasing the lines capabilities between Moscow and Western Europe, St.Petersburg and Finland. Now in the capital there are not less ten ways to get IP network access for acceptable money, although the prices are higher 2-4 times as much as in the US.

Contents

That person who may allows himself such daily wasteful amusement as two hours surfing over Internet differ from usual pioneer-networker. He does not understand programming more often than not. The customer has steady interest in electronic newspapers, computer games, arts, and, in addition has also trouble with English. So there will be information resources in Russian and they have been created. Traditional news groups with adverts for child's food and non-ferrous metals selling have ceded the palm to new Web pages with well-known newspapers «Izvestiya», «Computer World», Russian Internet guides, and notes about music, movies. However lots of these materials are accessible on subscription only. For example, you have an alternative either to pay $195 monthly and to read full texts of speeches on Moscow briefings or to do the same free but two weeks later.

One should note that in Russia there is academic network FREEnet (For Research, Education and Engineering) where you may not pay attention to traffic and connection time. This network has other goals in comparison with Relcom, Russia Online, and Sprint. It exists on meagre means of universities, city administrations, and grants of foreign foundations. FREEnet tries to share, first of all, educational and scientific information and is the main participant of a project named «Universities of Russia».

Technical details

The popularity of NCSA httpd in Russia has been reduced from 85% to 30% for last six months. The today's winner is Apache (45%). Perhaps simplicity of their installation and maintenance remains magnetic for beginning Web administrators. Next 10% work with CERN daemon, the others, they are rather isolated, use Netscape and Microsoft servers (10% for both of them). As you may note free software predominates.

Reasonable and well-founded choice of method of Cyrillic document presentation is a very significant moment for starting a new WWW site. Russia was an experimental ground of Microsoft that has invented even two different Cyrillic charsets: CP866 for DOS and CP1251 for Windows. Besides there are at least three other similar computer charsets used for Russian language support: KOI8 (or KOI8-r) - under a lot of Unix system; MacOS Cyrillic - technology from Apple Computer; ISO-8859-5 - the only Cyrillic charset supported in MIME and by large Unix manufactures such as Sun Microsystems, etc.

The most trivial and widespread solution is to create HTML documents in the frequently used KOI8 charset and to advise people everywhere on Internet to provide themselves with corresponding fonts. The more resourceful way is making changes in server code or development of a new HTTP daemon that serve text files in particular charset depending on TCP port or building several parallel hierarchies of files with choosing a suitable charset by hand. Some sites provide the transparent «on-the-fly» recording from server charset to browser one. In this case daemon determines a browser type and its charset through «HTTP_USER_AGENT» and/or «HTTP_ACCEPT» field passed by the browser to the server.

Prospects

The wave of WWW creation in Russia continues its running. Some computer magazines publish announcements about vacant positions for people having experience in HTML, CGI, Unix system, electronic document exchanging. It is an encouraging fact - level of presented materials is being improved due to new professional painters and information managers involved in this process.

At the same time accessible HTML documents are still old-fashioned and contain texts and GIF/JPEG graphics only. Unfortunately we could not find any example of state-of-the-art technologies: usage of Java, live and on-demand audio/video transfer. Perhaps it will happen in the very near future. We can expect also some reduction in the price of advertisement placement and various information services by increasing number of offers. This time is likely to be not so far.

Creating A Presence On The World Wide Web - Some Key Strategies

Elaine Lawrence
Lecturer, School of Computing Sciences
University of Technology Sydney
AUSTRALIA
email: elaine@socs.uts.edu.au

Robert Hudson
Web Design Consultant
Web Obsession, Tampa, Florida
USA
email: robi@allworld.com

Introduction

'Change is at the very root of business' [Hill 1993] and one change, with which businesses in the 90s are grappling, is the World Wide Web. Because the technology has a half life of about six months and usage is growing at about 30% per month, any organisation with a planning time line of a year is going to find itself left behind [Lanfear 1995]. This paper explores key strategies for organisations which want to create a presence on the World Wide Web. It draws on some principles of business and information technology management, including Total Quality Management principles, to provide a theoretical basis to the practical insights gained by the authors from their consulting practices, both in the United States and Australia.

Key Strategy 1 - Plan But Do Not Over Plan

Work involves planning, organising, controlling and monitoring activities [Hill 1993]. Creating a presence on the Web involves all of the above but with important variations. The traditional paradigm of gather data, evaluate alternatives and select plan, does not work with the World Wide Web [Lanfear 1995]. Companies need to adopt an evolutionary and flexible approach for setting up a web site. Because many companies have 1950s work philosophies they react to the World Wide Web by setting up large committees to discuss World Wide Web policies, content and implementation strategies. They often lose strategic advantage by taking too long to get a web site online. It is better to get some web experience and projects online quickly and constantly improve them instead of spending months arguing over policies and content. Rather than setting up large committees, the authors advise running strategic planning workshops which should be conducted by professional web consultants and should assist interested participants, including senior management, to become aware of the potential of the medium. It is vital to have a senior person on side if the web site is to get approval quickly. The first step in achieving quality excellence is the decision to make quality leadership a basic strategic goal [Hill 1993]. One half day could be spent training the participants in the World Wide Web and the next day working out strategies for how the company could create its presence on the web. Try to build quality into the page at this early design stage to conform with the principles of Total Quality Management.

Key Strategy 2 - Organise But Don't Over Organise

The authors believe that the web site should be thought of as a continually growing and informative picture of the company. As such it is essential that it is updated and constantly improved in quality levels as part of total quality management. There must be some hook to get people to return to the site otherwise it will not be revisited. Some companies are now adding in small advertisements for their and for others' products. It is also vital to organise training sessions for all members of the organisation who are interested in the web. The more people who are involved and enthused by the World Wide Web the better, as they will be able to provide new ideas for updating the company page. Training covering all aspects of server operations and web presentations should be provided [Lanfear 1995] because such good training will contribute significantly to standardisation and 'corporate image'. This idea is in line with company wide commitment and company wide introduction of Total Quality Management principles.

Key Strategy 3 - Control But Do Not Over Control

In many organisations control of the web site can become a power tool for a division or department. Office or organisational politics may cause serious conflicts over who has ultimate control. In some companies, divisions that have been downsized, may see control of the web site as a way back into favour in the organisation. Other divisions, for example computing services, may see the web design and maintenance as just another chore that they have to do on top of their already busy schedule. A SPIDER (Strategic Planning for Internet Dissemination, Evaluation and Retrieval) team can assist in such ventures [Lanfear 1995]. If different departments of an organisation want to add pages to an organisation's web site, they need support in such works but not complete control of every page that is added. Templates and guidelines can be circulated. Templates are useful as they assist the different departments to quickly add a page and it helps if they feel that they are not starting from scratch. The guidelines need to set out who has formal authority to approve pages. The USGS (United States Geographic Service) enforces such detail as having USGS identification on all pages and links to other USGS pages [Lanfear 1995].

Some large organisations, in particular universities, often find it difficult to exert control over web page creation as questions of academic freedom and stifling of creativity arise. One methodology that often works in such environments is insisting on design consistency down to the first two levels of pages in departmental sites and then allowing freedom of design further down.

Key Strategy 4 - Monitor Results But Do Not Over Monitor

Once an organisation has a web site it is a simple matter to gather statistics of hits on the site. Some companies favour visitor books or interactive forms. There is a danger of overkill on statistics. Web users tire of filling in large forms unless there is some incentive for doing so. It is possible to allow a demon to measure the number of hits. Each week, the service provider for Cyber.Consult emails the number of visits to the home page broken down into the countries of origin of the hits and the domain names. Obviously if a company needs more information interactive forms and visitor books will be necessary. In attempting to adhere to the principles of Total Quality Management, the gathering of information from the 'customers' who 'hit' the organisation's page can be used as a tool to ensure that the customer requirements are incorporated into the design and continuous development of the site [Hill 1993].

Conclusion

Creating a presence on the World Wide Web can prove to be a daunting task for an organisation, particularly if the organisation is a large one with management philosophies rooted in the 1950s. Companies need to take advantage of the dramatic developments that have taken place in communications. They must act quickly and decisively to put up a home page and explore ways to revitalise their business by taking advantage of the phenomenon of the web - they must not over plan, over organise, over control or over monitor the development of their web site but must take an evolutionary and flexible approach. If they adhere to the principles of Total Quality Management, they will find establishing a presence on the web an exciting way to promote their corporate image and do business in the future.

References

[Hill 1993] Hill, Terry (1993) *The Essence of Operations Management*, Hertfordshire UK: Prentice Hall.

[Lanfear 1995] Lanfear, Kenneth (1995) *Email note on Strategies for implementation of World Wide Web*, Networks Information Products Coordinator, Reston, Virginia.

Student reactions to a World Wide Web-Based Textbook: An Exploratory Study

Lisa M. Lehman
Assistant Professor of Information Science,
Rasmuson Library,
University of Alaska Fairbanks
Fairbanks, AK 99775, USA,
907-474-7403
lisal@muskox.alaska.edu

John A. Lehman,
Professor of Accounting and Information Systems,
Professor of International Business
School of Management,
University of Alaska Fairbanks
Fairbanks, AK 99775, USA,
907-474-6275
ffjal@aurora.alaska.edu

Abstract: The paper describes the use of World Wide Web-based textbook for a required MBA course covering business information sources offered in the School of Management of the University of Alaska Fairbanks. It raises theoretical questions about the which text models are most suitable for organization of such materials. Preliminary practical results are that students' reaction to the use of www-based materials is in part a function of currency, and in part a function of whether the materials were designed to take advantage of HTML, or whether they were simply translated from printed form. Preliminary theoretical results are that stream based and shallow hierarchical object based text models do not work well with the new medium. Unresolved questions revolve around which object model for hypertexts provides sufficient value to overcome the disadvantages of screen-based reading.

Background

This paper describes the use of a World Wide Web-based textbook for an MBA level course covering sources of business information. The course was designed to help these students find sources of information external to the organization and to become intelligent managers and consumers of such data. Specifically, the course focused on information outside the firm, whether in the form of library resources or on-line resources, and the management and presentation of data derived from these resources. The course was taught at the University of Alaska Fairbanks Graduate School of Management for four years. The course is that it was required for all MBA students at the University of Alaska Fairbanks. Students took the course after finishing the MBA core because functional knowledge of areas such as marketing and finance were prerequisites for the course. The information literacy provided by the course was used in other advanced courses in the program. It also served as the first integrative course within the MBA program.

The Textbook

The course made extensive use of electronic resources for course delivery and communication with the instructors. Because both instructors for this course had full time administrative positions and because many of the students likewise had full time jobs and were taking the course at night, communication had historically been a problem. E-mail, covered in the course, was used extensively for communication between the students and faculty as well as assignment distribution. This was particularly convenient for those students who traveled as a result of their jobs and who could therefore continue to work on assignments. Because the course was not widely taught, there were no suitable textbooks at the time covering the material outside those of traditional business research courses. In addition, organization of the Internet is politely referred to as chaotic and, therefore, the material available changed rapidly, even during the semester. We were forced to deal with a lack of textbooks, significant change in on-line systems from month to month, and a lag between data becoming available and that availability being documented. To overcome these problems, we created a World Wide Web based textbook.

The course's home page featured the syllabus with links from topics to related course material. The material included cases, readings, pointers to the on-line edition of NorthWestNet's *Internet Passport*, and assignments. The cases and readings were prepared by the two instructors. Access to the on-line edition of NorthWestNet's *Internet Passport* was provided as part of the University's institutional membership in the Northwest Academic Computing Consortium. Copyright on materials was therefore not an issue.

This version of the syllabus was designed as an HTML (Hyper-Text Markup Language) document using generally accepted guidelines for world-wide-web document design [DeRose 1994]. The on-line edition of NorthWestNet's *Internet Passport* used for the course, however, was an older version which essentially copied the text from the printed edition into files accessible from a web browser. It made no use of hyperlinks, and did not follow style guidelines for hypermedia. This provided us with an opportunity to explore the impact of both document form and content on user satisfaction.

The Survey

At the end of the course, we administered a survey gauging reactions to the Web based material. [Appendix A] 15 students responded. 14 had never used the Web before taking the class.

Comments indicated that the major problems were dialing into the university computer system and difficulties in reading on-line material. Computer-based text is not as portable (few students owned laptops) and convenient as a printed text. Many were accustomed to reading printed materials and had difficulty adjusting to screen-based text (vision problems were mentioned by several). Losing the continuity between pages and between links was often cited as a drawback. As one student commented, "... I think not having much experience with on-line materials could be the reason [for reading difficulties]."

The survey covered three issues:

1. whether the students preferred a web-based textbook to a paper one given identical contents
2. whether the students preferred a web-based textbook to a paper one given that the web-based one would be more up-to-date, and
3. whether the students preferred the syllabus, assignments, and other course material to be web-based or paper-based.

The results were that the majority of the students preferred a paper textbook to a web-based one given identical contents (10:5). However, given that the on-line version would be more up-to-date, the students preferred a web-based text (11:4). The students also preferred that the course materials be web-based (11:4). All of these preferences were statistically significant beyond the 0.001 level.

We suspect that the explanation for these differences are that the students did not find that there was much value added to the on-line text beyond currency. It had originally been designed and written as a paper document, and so neither followed principles of good HTML document design nor made significant use of hyperlinks. The syllabus however was designed as a web document from the start, and made significant use of hyperlinking, as well as following principles of HTML document design. These results lead to a number of interesting theoretical and empirical questions regarding the design of web-based documents.

Issues for Future Research

There are three basic structures for texts:

1. Stream
2. Hierarchical object based
3. Network object based

Stream structured texts are based on a typescript model and are similar to those prepared with most word processors. They are linear sequences of text data, consisting of a stream of words. Embedded markers indicated events (such as the beginning of a new line or a tab). The stream model can apply to non-text data as well; time-based media (sound, video) are almost always digitized as streams.

In object-based structures, textual objects such as paragraphs or lists actually exist rather than being represented by events as in the stream based model. For example, in HTML (an object-based markup system) each object is defined by a beginning of object marker (e.g. <BODY> and an end-of object marker (</BODY>). By contrast, a word processor (such as Microsoft Word™) marks a paragraph strictly by the end of paragraph event, and except for a few selection commands, does not deal with text objects per se. Anyone who has reviewed student papers (or conference proceedings) submitted in machine-readable form has experienced the wide range of formatting events used by different authors to define textual structures. On the other hand, there is only one way to define, for example, an unordered list in HTML, which treats such lists as objects rather than as a sequence of formatting commands.

Software for publishing large documents (such as Framemaker™) as well as markup systems developed for large documents (such as most document type definitions in SGML) usually adapt a hierarchical object based organization [Goldfarb 1990]. The hierarchical structure is traditional and the reasons for it are twofold. First, it has significant advantages in simplifying the human information processing problems arising from large structures [Lehman 1990]. Second, given the enforced linear structure of print, it is easy to implement. However, as any author who has ever force-fit a series of ideas into a hierarchical outline knows, the structure often does not fit the content.

Network object based structures attempt to deal with this problem by adapting an object-based model but not requiring a hierarchical structure. Such structures are difficult to implement in print (although footnotes and cross-references make the attempt); they are trivial in a computer-based document. They place additional demands on the user in terms of cognitive processing.

In terms of this theoretical differentiation of text types, the exploratory study described above offers evidence that users do not find that www-based implementation of the stream model and the shallow hierarchical object based model (book, chapter, section) add value to printed versions except in terms of currency. Unfortunately, the structure of our study did not allow us to distinguish between more deeply hierarchical object based structures and network object based structures. The question for designers of www-based educational and reference material is whether users find one or the other of these two structures better for the communication of information, and how a web-based hierarchical structure compares with a printed one in terms of user acceptance. We plan to test these questions in future studies.

Conclusion

As the Web's popularity has grown, the skills needed to find, read, and manipulate Hypertext documents will become more common. At the same time, it is important that authors of www-based material design and write them to take advantage of the new medium rather than mechanically translating from an older form.

References

DeRose, Steven and Durand, David, *Making Hypermedia Work*, Kluwer Academic Publishers, Boston 1994.

Hall, Hazel, "Setting the context for teaching business information," *Aslib Proceedings* vol. 48 no.1 (January 1996) pp. 9-15.

Goldfarb, Charles F., *The SGML Handbook*, Oxford University Press, 1990.

Lehman, John, *Software Design in the Fourth Generation*, John Wiley, 1990.

McKinnon, Sharon M. and William J. Bruns, Jr., *The Information Mosaic*, Boston: Harvard Business School Press, 1992.

Mintzberg, Henry, *The Nature of Managerial Work*, New York: Harper & Row, 1973.

Internet Domain Survey, January 1996. http://www.nw.com/zone/WWW/report.html

The CommerceNet/Nielsen Internet Demographic Survey.
http://www.nielsenmedia.com/whatsnew/execsum2.htm

Host Distribution by Top-Level Domain Name http://www.nw.com/zone/www/dist-byname.html

"The Online World of Steve Case", Business Week, April 15, 1996, pp.78-87.

APPENDIX A

The Survey Used

BA 691

Fall 1994

Please circle the appropriate response

Before taking this class:

1. Had you used electronic mail? Yes No

2. Had you used usenet? Yes No

3. Had you used electronic library catalogs? Yes No

4. Had you used gopher? Yes No

5. Had you used the world-wide-web (Mosaic, lynx,...)? Yes No

For the following questions, please mark the line with an arrow indicating your preference between the two alternatives, e.g.

hot cold
^

6. If both documents had the same content, would you prefer a textbook which is:

paper on-line

7. If the on-line version were up to date and the paper version were a year old, would you prefer a textbook which is:

paper on-line

BEST COPY AVAILABLE

8. What is your preference on the hypertext syllabus vs a traditional set of paper syllabus and paper assignments:

paper hypertext

9. What problems have you encountered with the on-line reading materials? (continue on back if needed)

BEST COPY AVAILABLE

Development of 3W graphical information systems

Dr. Frank Lochter, GIS-coordinator, GeoForschungsZentrum Potsdam, 14473 Potsdam , Telegrafenberg A51/C6, Germany, marco@gfz-potsdam.de

Sven Knauer, Application programmer, GeoForschungsZentrum Potsdam, 14473 Potsdam, Telegrafenberg A51/C6 , Germany, sven@gfz-potsdam.de

Problem

The GFZ is a nonuniversitarian, large-scale research centre in Germany. It combines all solid earth science disciplines including geodesy, geology, geophysics, mineralogy and geochemistry, forming a multidisciplinary research centre. The facilities include about 80 workstations and more than 300 PC's in the net. The scientific data are stored in Relational Database Management Systems (RDBMS) in a client server environment.

Our information resources (geoscientific data) often include spatial components. Tools being developed to hold and search this kind of information. They must serve the following purposes:

- interactive graphical spatial retrieval
- interoperable, running on microcomputers and workstations
- access to distributed data sources
- visualization of data stored in relational database systems (Sybase)
- central administration of the tools

Solution

The data centre of the GeoForschungsZentrum in Potsdam is tasked with the development of an infrastructure for multidiciplinary investigations of the lithosphere. A metadata system "GEOLIS-GFZ" {1} has been implemented in order to store (in RDBMS) and find the data and information resources. It is an information system about information systems. Standard metadata formats like DIF (Directory Interchange Format) are used for the input of metadata.

At the data centre of the GeoForschungsZentrum in Potsdam the functionality of WWW-Servers with CGI (Common Gateway Interface) has been extended via PERL-programs to implement graphical information systems. The following applications have been developed:

The "GFZ-Spatial Retrieval" {2} (Snapshot) application is able to retrieve data stored in the meta information system "GEOLIS-GFZ". The user can search for objects specifying thematic, spatial and temporal attributes. We've implemented the possibility to specify a search rectangle on a map background for the spatial retrieval. A geo-thesaurus (predefined regions of the world with the corresponding bounding rectangles) helps to find the region of interest. A thematic thesaurus has been implemented as source for the thematic attributes, but the specification of free terms is also possible. First information systems using this functionality are the GFZ-Map-Library and the management of limnological data.

The "Knowledge Browser" {3} (Snapshot) is a tool to build and explore information spaces. Until now it's only a prototypical solution demonstrating the retrieval of alphanumeric documents by using a Virtual Reality based retrieval interface. The "Knowledge Browser" allows the navigation in virtual information spaces and the interactive selection of objectes in these spaces.

The "GIS-Glossary" {4} (Snapshot) is a tool to explain GIS-terms. The HTML-pages will be produced "on the fly". They include the search term, it's explanation and a picture. The tool can be used for any kind of glossary.

URL's

- {1} http://www.gfz-potsdam.de/drz/geolis
- {2} http://www.gfz-potsdam.de/cgi-bin/spatial_retrieval
- {3} http://www.gfz-potsdam.de/cgi-bin/knowledge_browser
- {4} http://www.gfz-potsdam.de/cgi-bin/geolis/general/glossary

---

Developing Technological and Content Knowledge through Publishing Web Pages

Carol A. Long
Department of Special Education
Winona State University, U.S.A.
clong@vax2.winona.msus.edu

Sue P. Reehm
Department of Curriculum and Instruction
Eastern Kentucky University, U.S.A.
elereehm@acs.eku.edu

Mary Ito Dennison
Department of Special Education
Winona State University, U.S.A.
mdennison@vax2.winona.msus.edu

To enhance learning and the acquisition of technological skills, a home page was developed and made available on the World Wide Web. Preservice teachers at Winona State University enrolled in a Special Education course on characteristics of students with mild disabilities were required to develop interactive web pages connecting content, recent research, and local, state, national and on-line resources pertaining to learning disabilities and mental retardation.

The project was done in four phases. During the first phase information was gathered and plans were completed. Plans included how to organize the students into teams, what technological training was needed, and how to assess computer and inter net competence and attitudes. Students were introduced to the project in the second phase and were randomly assigned to teams with graduate students designated as team leaders. One group was charged with creating a web page on the topic of learning disabilities and the other group was required to generate a web page on mental retardation. During the second week of class, groups were organized by subdividing into smaller teams and topics for subsequent web pages were assigned so that information gathering could begin immediately. The third phase consisted of directing the students as they generated the web pages. The graduate students were instructed in HTML and web page design. The undergraduate students were assisted in researching information for page content on learning disabilities and mental retardation.

The students created 15 web pages which were linked to the Department of Special Education home page. The pages provide definitions for the two disability categories which reflect national and state interpretations. Listings of common characteristics of children with learning disabilities and children with mild mental retardation are given. The students included a bibliography of recent research on learning disabilities and mental retardation and resources for locating further information and assistance.

The fourth and final phase of the project was the evaluation phase and was done in three parts. During the first phase, a pretest was administered to determine attitudes toward technology and perceived Internet competence. In comparing the results of the questions, the students felt it was very important for teachers to know how to use a computer but ranked their knowledge of computers below the computer knowledge of other college students. The students reported the need for computers to be used in every classroom and most reported they were comfortable working with computers. T-tests were used to compare the means for the graduates and undergraduates on each question. There was no significant difference found between t he two groups on any other the questions.

At the completion of the project, the same survey was given as a posttest. No significant difference between the pre- and posttest was found on the questions related to the attitudes and general use of technology. However, a comparison of the results on the pre- and posttest Internet questions found several of the scores were significantly different. Analysis found significant increases in the students ' knowledge levels for locating information on the Internet, using Netscape, using Gopher, defining search engines, defining home page, using Telnet, defining URL s, publishing home pages, listservs, e-mail, and FAQs. None of this information related to skills directly taught during the course. The students learned how to use the Internet by locating information on the web and gathering and organizing the information to create the web pages.

In the posttest, the students were asked to compare their knowledge gained from this project to projects in other classes. The students rated their learning greater from this project in the areas of content knowledge, working as a

team, and programming than projects in other classes. In both pre- and posttest surveys, students acknowledged the importance of technology for teachers but were uncertain about the level of their technology skills.

Several conclusions can be drawn for the analysis of the data. The students involved in the project acknowledged the importance of technology for teachers. However, they were uncertain about the level of their technology skills. This was true for undergraduates as well as graduate students and before and after the project was completed.

After the project, the students reported an increase in their perceived abilities to use various Internet tools. They felt the project was worthwhile with graduates rating it more so than undergraduates. Students thought the project was superior in several ways to projects assigned for other classes but that it required more effort on their part. Comments included "introduced the class to technology t hat exists and is growing," "I have never done something like this before-exciting but lots of cooperation and hard work required," "I learned not to be afraid to try anything," "What a great accomplishment for WSU Special Education. Way to be a leader," "I am computer illiterate and this gave me a new look at how computer s can be used," and "I learned so much from it-It was an exciting experience."

IDEALS, An Approach to Web Based Remote Training

J. M. Brisson Lopes(*), José Carlos Teixeira(**)
(*) IST - Av. Rovisco Pais, 1096 Lisboa Codex, Portugal
Email: brisson@malhoa.ist.utl.pt
(**) CCG/ZGDV - FCTUC - R. de Moçambique, 17 R/C Esq. - 3030 Coimbra, Portugal
Email: teixeira@ccg.uc.pt

Introduction

The Web has dramatically changed the way we used to look at the world. The emergence of the Web and Web viewers has made world-wide available an almost infinite number of information sources with such an extent that no one can grasp anymore. Use of the Web for educational purposes is one of the many fields of application that came naturally to be. The Web presented itself as the ideal environment to deploy Computer Based Training (CBT) applications and services. The Web's support of most multimedia formats looked extremely promising for learning and authoring. If courseware could be created with high quality multimedia educational contents, one would inevitably get the best rated courseware. Moreover, Web widespread usage would ease learning and training by putting high quality courseware at the fingertips of anyone person regardless of geographical location.

A very common belief is that no one teaches but everyone learns. Conventional CBT systems gallantly walked into this false illusion. Successful learning can not rely on the superiority of the learning material alone. Person to person contact is also a must. Learners must keep some contact with teachers and other learners taking the same courses. Today, the Web has made it possible for people to contact other people live. It is only necessary that the Web integrates the latest developments in teleconferencing. As a consequence, CBT systems must evolve and adapt to the new needs. This calls for a new generation of CBT systems which is aware of both the Web and all the psychological and educational aspects of learning to accommodate the predictable scenario changes that will start to take place very soon.

This paper addresses the likely changes in learning scenarios and user needs and the role that the Web can play in this context. Development of advanced, Web aware CBT systems has already begun. One of such advanced systems is IDEALS which will be described in this paper. IDEALS aims at creating and deploying advanced CBT software and services based on the Web and aware of the specifics of learning.

Trends in learning scenarios

The two driving forces of the global society are the need to adapt to continuous technological change and the simultaneous demand for better qualified person on the labour market. Retraining on the job has a vital need for most firms, especially small and middle sized firms. The emergence of learning services fits into the space opened up by the new learning scenarios that present the following demands:

- universal, comfortable and equitable access to information

- information diversity and choice

- use of open and interactive networks, at a reasonable cost

New learning and training scenarios should support the development of new skills and enable the creation of new kinds of jobs through innovation.

Lifelong learning is one of the key components of a new global economy, where knowledge is the critical resource. One of the most important goals of someone involved in learning and training is the identification of the best enabling technology to use. The use of distributed based learning scenarios opens an opportunity to harmonize contents and methodologies and, therefore, learners skills, increasing the mobility effectiveness of professionals, both at national and international level.

The Web Based Solution

Most CBT applications remain monolithic long after the emergence of client-server technology. The emergence of the Web opened up a complete new universe of opportunities for CBT. First, it no longer matters which platform learners use for learning since standard, off-the-shelf Web clients are available at reduced or no cost. Once the Web's client software is installed, users can immediately start learning. Second, the difference between local and distance learning was significantly shortened by the Web and related communication facilities. On the server side there is also an enormous potential to explore.

Moreover, the rapid evolution of Web application tools is already providing most of the tools needed to build next generation Distance Learning CBT systems capable of satisfying the requirements of tomorrow's learning scenarios. Issues like security and accounting are already finding applicable answers from tools that are less than one year old. On another side, the Web makes it possible to build courseware co-operatively with geographically separated authors contributing in real time in courseware development. We may say that the Web and its developments have created the ideal environment for the development and deployment of the next generation of CBT systems.

The IDEALS Approach

IDEALS is a project of the Education strand of the Telematics Programme sponsored by the European Union. It targets at producing and demonstrating an advanced, Web aware Distance Learning CBT system that meets the training needs of organizations and universities in tomorrow's learning scenarios. Project demonstration will take place in Small and Medium Enterprises and Universities across Europe. Courseware developed with the IDEALS CBT system will be delivered to trainees at the workplace, home and learning institutions.

IDEALS CBT system, the Modular Training System (MTS) integrates Web tools and other learning oriented tools developed within IDEALS. MTS is a server that supports all learning oriented operations requested to it by a normal Web server.

Authors selected from the best experts in each field are contributing to build courseware. Each contribution is locally stored at the site of the contributing author and linked to the other contributions to make up a complete course. All sites are sites of a common distributed database and links between courseware material parts can be virtual links. Authors need not know the exact reference to any other module (e.g., URL). Courseware modules posses a characterization that is stored together with each modules in the database.

Since IDEALS MTS assumes the fact that courses are build from modular courseware material, it allows the creation of complete new courses on the fly, from the material already present in the database. All the course author needs to do is to create a higher level course module, called ***course node***, which refers lower level course modules, the ***course units***. Course nodes list and specify the sequence of modules in a course. Of course, the references to other modules are virtual.

At course runtime, IDEALS MTS will inquire the distributed courseware database to find the course units that best fit virtual references specified by authors in course nodes. This operation is IDEALS Mapping. IDEALS MTS takes into account that learning is a personal activity and, consequently, each person has his/hers personal learning profile, speed and performance. IDEALS design had all the time in mind the notion of constant adaptation to the user. MTS measures and records learning speed and success as well as other factors (e.g., learner background knowledge). Virtual references are only mapped when so needed, never in advance. Therefore, MTS has the possibility to adapt course flow to actual learner state. With IDEALS MTS the two above learners would still study the same subjects but would proceed through different course units selected during runtime according to each person's specific learning profile and degree of success in successive course steps.

IDEALS evolved from early experiments and research projects as DEDICATED project, which was an extension of the classic electronic classroom located in a training center where to learners needed to go. The concept of locations for training purposes (Local Training Centers) was extended by interconnecting them and effectively real-time pooling and sharing their courseware. All sites now form a Virtual Training Center. Learners need no longer go to any LTC to learn and train. They can now stay at their workplaces or at home and learn whenever they choose, for the length of time they want.

A Virtual Campus on the World Wide Web

Rory McGreal
TeleEducation NB
Department of Advanced Education and Labour
New Brunswick CANADA
Email: rmcgreal@unb.ca

The New Brunswick Community College (NBCC) is a province-wide system of 10 campuses in all regions. The College is now establishing an additional Virtual Campus on the World Wide with courses and programs available for both on and off campus students. The Virtual Campus consists of four units responsible for the following activities: content development, service delivery, export, and research and development. New tools and techniques are becoming available to teachers and learners on the WWW which allow them to communicate do research, and interact on line in ways that have not been feasible in the past. The NBCC sees the new Virtual Campus as a means not only of reaching students in isolated regions through distance education, but also as a means of providing better quality training to students on the campuses and in workplaces.

Current thinking in telecommunications is paralleled in other human organizations by a reliance on hierarchical structures. However, in telecommunications as in human organizations, world views based on the limitations of copper wire and a shortage of bandwidth and information that must be controlled are fast becoming redundant. This change in thinking is predicated on the rather rapid development of networks where intelligence is distributed around the periphery, rather than at a large control center. Each and every node on the network, and each and every user of the network can now publish text and images, function as a radio station and shortly will be able to open a TV studio. The World Wide Web (WWW) opens up this new distributed environment for use by everyone. It can also be adapted for learning and training from anywhere to anyone, anywhere.

Computer professionals are supporting more freedom with less control from central authorities. The Internet is an example of a free and open network. Telcos prefer to control and manage, and their networks function in a controlled environment. The NBCC task is to determine the right mix between anarchic freedom for the individual campuses and full control from the center. An open and distributed network allows free access with no center of control. Before implementing policies and initiatives to take advantage of the environment, the center and the peripheries must collaborate more in determining the optimum mix of central guidance and coordination while maximizing the benefits of local control and independence.

Virtual Campus Structure

The Virtual Campus structure will include four working units distributed around the province with coordination from the center in Fredericton. The units will be responsible for content development and maintenance, service delivery, export and marketing, and research and development. The content development unit will be responsible for adapting courses and integrating new learning designs and media into the course delivery. Each campus can participate. Training, coordination and help from instructional designers, graphic artists and programmers will be made available on line sessions, workshops and tutorials, and through on site visits by specialists.

The service delivery unit will be aided by the existing TeleEducation NB network of over 80 sites in more than 46 communities.. In collaboration with Community Access Canada, local schools, libraries and private sector partners, new sites are being added giving all citizens in every community at least one point of access to the information highway. The network of eighteen regional site facilitators, local volunteers and technical support personnel are responsible for ensuring that the network is accessible and user friendly at the local level.

The export and marketing unit will support the development, packaging, and marketing of educational products, programs and other services such as consulting and franchising. The research and development unit will work in close collaboration with provincial universities, the TeleLearning Research Network and CANARIE to identify needed, practical and feasible research projects.

In addition, other provincial and out-of-province institutions in both the public and private sector will be participating in using the virtual campus environment. Templates and other online tools as well as the online environment will be made accessible for development and delivery purposes.

Virtual Campus Activities

Programs offered by the Virtual Campus will take full advantage of the on-line features available, as well as relying on more traditional media like print, fax etc. Traditional college programs contain the following elements: content delivery, interaction with teachers and students, skills development, and practice. Face-to-face lecture is the principal mode for delivery of content. Asking questions is the main form of student-teacher interaction. Skills development is hands-on, as is any meaningful practice by the students. The WWW, combined with on line access to Computer Based Training courseware, a workbook, and a textbook can provide a more effective platform for delivering program content. The online environment can also be used to facilitate student teacher interactions, and in some programs provide hands-on experience and opportunities for skill practice.

The following activities are presently being used by learners on the WWW:

1. Listen to, read, or watch a video of lectures
2. Ask question of the teacher and of individual students
3. Participate in certain experiments or simulation of experiments
4. Deliver student presentations.
5. Provide access to online courseware
6. On line student support programs
7. Library searches, and journal browsing
8. Work on projects with other students
9. Undertake statistical analyses
10. Social chitchat
11. Group discussions among students

Web software is now becoming available to support all these activities. The NBCC can also make effective use of existing software, hardware and telecommunications environments. For example:

1. Gopher, Telnet and FTP; (now possible through the Web browser)
2. Email and Usenet groups, chat lines and keyboard conferencing;
3. An 800 number for just-in-time individual aid to students;
4. Courseware made available on -line and downloadable;
5. Remote support using software for screen sharing, audio and video conferencing.

Virtual Campus Features

This new online learning environment can be used effectively for the delivery of nearly all content materials. This has the additional benefit of transferring much of the responsibility for learning from the teacher to the students. Employers faced with a rapidly changing working environment are increasingly demanding the skill of 'knowing how to learn' from their employees. In addition, teachers liberated from lecturing can more effectively concentrate on giving students more individual attention and spend more time on skill training and monitoring.

The Virtual Campus is distributed. All physical campuses will be available for delivery and reception of courses as will all TeleEducation NB sites. Staff at all campuses will have the opportunity to participate in the development and delivery of programmes, and will be expected to facilitate. The Virtual Campus must work closely together with the different campuses, TeleEducation NB (the delivery network), and the Center to determine the best mix of centralized coordination and campus independence. The goal in implementing policies and initiatives will be to take advantage of the on-line environment to determine the optimum mix of central guidance and coordination while maximizing the benefits of local control and independence.

A unified NBCC-wide (and hopefully integrated with all the provincial secondary schools and universities) student records system including all campuses will be a key component of the on-line initiative. This system will include facilities for on-line registrations and payment of fees. Student loan information and applications will also be available, with the option for students of completing the forms on-line. A system-wide student support network including guidance, special help and other services, will be readily accessible to all students whether on campus or at a distance. A competency-based testing system is being implemented in order to ensure standards across the system and also to ensure that students are trained to internationally acceptable standards in whatever field they are pursuing. A common system for accreditation, and recognition of life experience and outside credits has been developed. This system will address the articulation of courses between universities and the College, and between the schools and the College. An approach that encourages collaboration with the private sector and other colleges outside the province for the effective sharing of resources is being fostered.

Virtual Campus Programs

The Virtual Campus will take advantage of other available technologies. Many programs may be integrated with on-site programs. Full-time and part-time students would be using the Virtual Campus for some proportion of their studies. The first programs now being developed for delivery through the Virtual Campus are based on new courses in Multimedia, and in French: Bureautique. The Virtual Campus is focusing on specific programs like these that are identified as being more amenable to on-line delivery. In collaboration with four other provinces, a nation-wide course in Information Technology is now being developed. Each province will create one module that will be shared by all the participants. It will be available in September 1996 in French and English. This will be followed up with the development of other courses, leading to a full program, following from this initial introductory course.

In addition, there is an acute need for staff training, now that the Virtual Campus infrastructure is being implemented.. There are individual teachers and other staff in all campuses that are familiar with the new environment. The challenge is to build on this base and expand out so all teachers will feel comfortable when navigating in cyberspace. After the installation of the telecommunications and software infrastructure, effective staff training is the key element in the building of the Virtual Campus. The programs that we develop can be doubly used to train teachers. Moreover, those teachers who are involved in the development of online programs will form the cadre of experienced cyber professionals that are needed to build the Virtual Campus and transform the provincial community college system.

Conclusion

When the NBCC makes programs available on the World Wide Web, they will be accessible from anywhere in the world. New Brunswick students will be accessing courses on line whether the Virtual Campus exists or not. Students will be able to choose these programs from all over the world. New Brunswick has a technological lead over other jurisdictions in the physical telecommunication infrastructure, based on the penetration of digital optical fibre laid down by NBTel to every community in the province, and access to more than 93% of homes through Fundy Cable. Moreover, the province is small enough to be able to move quickly.

Initiating change in large organizations can be problematic. Bureaucracies are stable systems, and in normal times this is considered very positive. However, in times of great change, this stability can prove to be an insurmountable obstacle to bringing in required changes. One way of introducing change is by adding on a process or system rather than trying to force change within. By focusing resources on the new system, the change agents can establish a beachhead within the main system that can then be used to promote change in the main organism. New processes or systems that are unknown to those within the system often have a better chance of success, because they are not seen as threatening.

The WWW can be used to implement new NBCC programmes that are not threatening to the organization. By focusing on new programmes, we can also use them as test beds to help us in the design of the new environment. These test beds can be used to demonstrate the viability and the limitations of the environment. The lessons learned from the Virtual Campus initiative can then be used when we broaden the environment to encompass other programs and move toward full integration of the online courses into the traditional campus programs.

Information Technology and the Viking Network

William J. Meador, BS, MA;
Victor Elem. Sch. Dist.
Victorville, California
edsearch@primenet.com

Why would culturally diverse classrooms of students be interested in what occur-red, a thousand years ago, to people who spoke a foreign language and ran around in skins and metal skull caps for helmets? They could be forced to study the Vikings as part of a national and/or local curriculum, but how much would they learn or retain and of what importance is it to their future? Much more is gained if the goal is teach IT from a historical perspective as the tool. As educators we are all seeking the most advantageous and motivational methods and devices in which to keep our students as active learners and producers of knowledge. For many of us IT is the thrilling and ever expanding tool for our students, and ourselves. We, as teachers, also need the challenge to continue to be active learners.

The Viking Network is the answer for some educators in eleven different countries including the United States. Since 1993, when conceived, it has continued to grow. Not only are the students dictating or writing messages to students around the world. They are searching for data and information about the countries they contact and building web pages to "boot" for easier access to one another. It is a living and growing entity. The students are now the collaborators for research projects, the innovators of thought projection, and the authors of historically based fiction. We are about as sure as we can be that the Vikings never made any forays into the Southwestern deserts of the U.S., but we can imagine, think, and write about the possibilities with their having a chance meeting with the Navajos or any of the Native Indian populations evident at that time in history. How about the Vikings trading with the Chinese? For those students who have a flair for science fiction perhaps an encounter with aliens.

The Vnet is adding to its projects to reach most of its students. We all have unmotivated students or those whom have less than average ability levels. These are the students that present the greatest challenge to me. This year will see an increased interest in IT by most of our students. Many of us on the Vnet see ourselves as teachers for all of the students in the Vnet. We have the feeling of a local network that just happens to be international and wish to be regarded as an expanded or open classroom.

This Fall is an exciting time for the members of the Viking network. Some of the projects planned for this year include:

1. Collaborative problem solving across the curriculum;
2. Mathematical quizzes and contests between students and classrooms;
3. Discussions about the Viking influence on the English Language (The Dane law Pages);
4. Student built web: Local activities, holidays, and celebrations presented enhance the learning of history. Local history gives an international dimension and audience information available on the children's very own web pages. These pages will be very personal to each student and his/her needs as a thinking, expressing and learning person.

An example of a very successful interactive exercise took place late last year between two classroom in separate European countries. The setting was that of a monastery where priests resided. The priests (classroom A) were notified of the impending arrival of a band of Vikings (classroom B) desirous of capturing and occupying the Monastery. Each classroom planned their strategy pertinent to their position and indicative of their goal within the game. The game lasted several weeks to the lasting enjoyment of the students and teachers involved. The Vnet gives NEW opportunities in teaching because information given from each local school is local information not readily available. A subject area is studied from several angles not from information from one's own country's writers and historians.

Teachers will interact with all of the students and each other via their own web page with posting of messages, queries, and responses to students and teachers alike.

These projects engage our students to be knowledge producers rather than just consumers.

Start Your Search Engines! Authentic Activities and the Internet

Eunice M. Merideth, Ph.D.
Associate Professor and Chair, Department of Teaching and Learning, Drake University, USA
E-Mail: EM1661R@Acad.Drake.edu

Hilda L. Williams, Ph.D.
Professor Department of Teaching and Learning, Drake University, USA
E-Mail: HW9721R@Acad.Drake.edu

Industrialization and modernization on a national scale became possible only when humans were able to harness natural sources of power--wind, water, fossil fuels, the atom--to release energy. The abundance of such energy created force, allowing the systematic organization and operation of work. The harness applied to nature's power was that of the engine, "a machine that changes energy into mechanical motion" (Compton's Interactive Encyclopedia, 1996). In today's Information Age, we are facing a revolution of another kind--the explosion of information and communication via the Internet--for which we need new types of engines that will do work, save human energy, and focus power. Fortunately, serious Web users have a variety of engines, appropriately called search engines, to help them do just that.

This paper presents a Three Step Approach to Integrating Web Applications and Services, harnessing the educational potential of the Web through the use of authentic activities (Brown, Collins, and Duguid, 1989) and generative learning projects (Ryder & Hughes, 1997). *Step I: Critical Thinking Assignments* which require analysis and comparison of search engines and intelligent agents in terms of usefulness for educational informational retrieval as students learn cooperative problem solving to discover potential resources, develop research skills, and create their own meaningful subject-based linkages to the Web. *Step II: Authentic Activity Procedures* which accept learners as thinkers with valid questions and the ability to create meaning and materials. These procedures would be rooted in authentic activities which result in observable performance. *Step III: Published Products* which share results of investigations or add original materials and resources to an evolving knowledge base available on a Web site. This three-step approach can be adapted to either undergraduate or graduate higher education curricula as described below.

Undergraduate Curriculum

Following an introductory lesson reviewing basic definitions and navigational strategies using a Web browser, we distribute a handout with the address and brief annotation of a number of search engines. We ask students to use one of the search engines to seek resources on their curricular topic in preparation for lesson planning and presentation.

Step I--Critical Thinking Assignment: Find and download resources from at least four Web sites that will aid in the development of lesson activities in your subject area that you will be willing to share with other professionals on the course Web site. This assignment for undergraduates reinforces their new skills in navigating and retrieving from the Web even as it helps students understand that quantity does not equal quality, and information does not equal insight.

Step II--Authentic Activity Procedures: Constructing meaning involves evaluating information by sorting and sifting, saving only what is important. This authentic activity builds on a process most of us learned as part of our first research paper assignment--indexing cards. With the help of technology, we can take that process and update it by having students build their own database on a hard drive or disk, directly cutting and pasting information from the Web to the database instead of downloading files indiscriminately. The use of a relational database allows flexibility for further use of resources and addresses. The development of lesson activities using the database makes it personally relevant and a professional resource that one could continue to expand.

Step III--Published Products: Performance testing of an authentic activity should be non algorithmic, complex, and involve judgment and interpretation (Resnick, 1987). By publishing the product of their searches, undergraduates are authentically challenged to use Web resources to create a new resource which is then available for analysis and use by other professionals in the field.

Graduate Curriculum

Students enrolled in an introductory research methods course complete assignments designed to engage them in activities that reinforce the knowledge and skills associated with scholarly inquiry, including productive search strategies, identification and critical analysis of resources, synthesis of relevant information and finally, developing a report for review by other scholars in the community of learners.

Step I-- Critical Thinking Assignment: A major assignment for this introductory research methods course is the completion of a literature review on a topic of educational interest using both traditional print materials and the resources published on the Internet. As a first step, students complete a comparative analysis of the search engines noting the search strategies, number of hits, percent of relevant hits and overall strengths and weaknesses of each engine as relevant to their own research topics. They then present a report of their findings to other members of the class, compare their results and conclusions, and construct plausible explanations for discrepancies in their findings.

Step II--Authentic Activity Procedures: Students build their own database on a hard drive or disk, directly cutting and pasting information from the Internet to a relational database instead of downloading files indiscriminately or taking copious notes. Students evaluating resources gleaned from the Internet are engaging in the scholarly activities which they are studying.

Step III--Published Products: The product which students develop to demonstrate their having learned how to use the search engines is referred to as a *Topical Internet Tour*. Students find specific related sites, make bookmarks for productive sources, extract the bookmarks, and save the resulting file as a clearly defined sequence. They record only those links that prove productive for examining their topics as they sift through the myriad of information. This results in a source file consisting of a unique organization of resources and a permanent record for future reference. Documented tours serve as models for other researchers and promote a positive atmosphere for collaborative and cooperative exploration and learning.

Summary

Learners with a dynamic need for information can find a plethora of resources through search engines on the Web. But learners also have a need for a process which enables them to synthesize and organize that information so that it is useful. Like programming an automatic assembly line driven by mechanical engines, a human must be ready to make decisions about quality and quantity of information desired and act on those decisions. The Three-Step Approach to Integrating Web Applications and Services provides proactive procedures for students to search, sort, save, synthesize, create, and give back to an electronic community of learners who continue to weave the Web.

References

Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher 18*, 32-42.

Compton's interactive encyclopedia. (1995). Carlsbad, CA: Compton's NewMedia

Resnick, L. B. (1987). *Education and learning to think*. Washington, DC: National Academy Press.

Ryder, R. J., & Hughes, T. (1997). *Internet for educators*. Columbus, OH: Prentice Hall.

Observations on Web-Based Course Development and Delivery

T. Craig Montgomerie
Division of Technology in Education
3-104 Education North
University of Alberta
Edmonton, Alberta
Canada T6G 2G5
Phone: (403) 492-3667
Fax: (403) 492-3179
E-Mail: Craig.Montgomerie@ualberta.ca

Dwayne Harapnuik, MLS
Division of Technology in Education
3-104 Education North
University of Alberta
Edmonton, Alberta
Canada T6G 2G5
Phone: (403) 492-3667
Fax: (403) 492-3179
E-Mail: Dwayne@oanet.com

This paper discusses the development and delivery of a credit course: *The Internet: Communicating, Accessing & Providing Information* [Montgomerie & Harapnuik 1996] which is delivered completely over the Internet. The purpose of this course is to help students understand what the Internet is (and isnít), and learn to use Internet tools. Specifically this course prepares students to operate effectively in the knowledge society, to learn to use electronic communications, to access electronic resources, to prepare electronic resources for the Internet, and to understand and examine critical issues related to electronic communications (Privacy, Security, Copyright, Censorship, etc.).

While still in development (June - August, 1995), it became obvious to the authors that as proponents of alternative methods of instruction, and as purported ìexpertsî on the use of the Internet in Education and Library and information science, we should ìwalk the talkî and develop the course in such a way that it could be delivered completely over the Internet in an asynchronous mode. Equally, it became obvious that we really didnít know enough about a number of issues (e.g., the kinds of problems that students would encounter and the questions they would ask) to develop the course directly for Web delivery. It was decided that we would deliver the course in a traditional ìFace to Faceî (F2F) manner for one year, then redesign it for Web-based delivery the next.

The course was delivered for the first time in a F2F mode September-December 1995. A number of Web pages were developed to support this delivery. The course was offered a second time January - April 1996, again in a F2F mode. While the course was being delivered students were asked to provide information on what would make the course more amenable to use by the distanced student; this input resulted in the constant revision of the Web pages. During the second offering of the course, a few students who could not attend the lectures for a number of reasons were encouraged to still take the course and to rely on the new pages, but heavily encouraged to communicate with the instructors by telephone or electronic mail if they had questions.

The course was delivered completely over the Internet for the first time over the period May-August, 1996 with over 100 students enrolled in the course.

Course Design

The course was designed to follow good adult learning principles, particularly that the user should be in control of their own learning (content, pacing, and sequencing), that alternative methods of learning the same material should be available, and that the subject area for assignments should, if possible, be the studentís choice. We also tried to emulate all the different kinds of support that traditional F2F students have. The introduction of Netscape Frames gave us the ability to make all aspects of the course available to the student on the same screen. By implementing a ìside frame,"ì ìledge framesî and ìmain frame,î we were able to introduce our students to the navigation icons that represent course modules, help, resources, forums, evaluation and all other aspect of the course. Students can click on any navigation icon to load the desired page in the main frame, and still have all other components of the course immediately available on the other frames. Following a great deal of revision, the home page for the course has stabilized as shown in Figure 1. It is extremely important to expose a learner to the most important features of an on-line course as quickly as possible. We introduce the most important icons in the main frame and the navigation icon is presented within the first paragraph of the introduction screen.

The icons in the ledge frame represent the most often used components of the course: FAQ (Frequently Asked Questions), Help (which gives access to discussions of different problems, to the Teaching Assistant and Instructors), Navigation (which explains how to move through the course), Registration (course availability and how to register), Instructor information, Marks (including a summary of the marking criteria for each assignment and individual controlled access to each studentís own marks), Conferences (which links to the discussion groups set up to discuss issues and get help) and News (current course news and announcements).

The side frame gives the student access to the actual instructional modules of the course and their related reference resources. A toggle button at the top of the frame gives the student option of switching from the links to the course modules to likes to each modules reference resources. With the Course Module button activated, students have access to the instructional and assignment modules. The Introduction module lays out the format for the rest of the modules in the course and also provides the student an overview of what is required. Each subsequent module deals with aspects of the Internet and on-line communications that students must master to effective participate and contribute to the Internet. When the Course Information and Resources button is activates students have access to the course reference material and resources. All reference material is available on-line. Each of these sections point to the most current and effective resources available on the Internet.

One of the more challenging aspects of maintaining this course is keeping track of the changing resource URLs and updating the list with newer and more effective material. Other than internal links that point to components of locally produced course resources, most links point to sites throughout the world. One exciting part of putting together an on-line course is that instead of telling the student where to go to get material or information you can actually give that student a link the actual resource or expert. Immediate access to experts and their material make the Internet a very powerful instructional tool. For example, rather than designing an HTML tutorial, we pointed (with permission) our students at Alan Levine's *Writing HTML: a tutorial for creating WWW pages* [Levine 1995]. This has been rated by many as the ìbestî HTML tutorial on the Web. We concur with this evaluation, and since permission has been granted for our students to use it, it would be a tremendous waste of resources to attempt to replicate it.

As the course progressed, it became obvious that, since most of our students were connected via modem, the speed of page loading was a priority. Many ìnice, but not necessaryî graphics were deleted, backgrounds were simplified and a more efficient conferencing system was sought in order to improve the speed of the course.

Student Learning

The lack of classroom interaction and discussion is the greatest determent to distance learning. We are social beings; we need to interact. An enormous amount of learning takes place in through peer interaction. Until desktop videoconferencing technologies are fully mature and bandwidth on the Internet increases our best option for simulating or providing interaction between all participants are computer conferencing systems. The NetHowTo Forum became our direct link to the students and their link to the instructors. Equally important, the NetHowTo Forum became a link between students. To facilitate participation, students were required to participate in two different public forums or conferences dealing with two different issues. To earn 10% of their final mark each student was expected to initiate at least two new discussion points and was required to respond to at least four other discussion points as well as responding to the discussion points that they initiated. In addition to the issues discussion, another 10% of the final mark was based on contributions made in the various Help Forums.

One of the most interesting aspects of the Help Forum was the support and encouragement that students offered to each other. At the senior undergraduate and graduate level it is not uncommon for students to be fiercely competitive. This was not the case in the Help Forum. Instead of keeping hard earned knowledge students

Figure 1: EDPSY 497/597 WWW Page

consistently said things like: "I understand the frustration, I was frustrated too the first time I....î or I was overwhelmed too but I soon found that....". We are finding that there is a high level of student cooperation and encouragement in this course. Ironically, the distance between students is one common factor that is bringing them together on line. The issues discussions are incredibly interesting and rewarding to follow. Some examples of the many good discussions can be accessed at: http://www.quasar.ualberta.ca/nethowto/articles/forum.

The scope of the final assignment is very broad. Students are asked to construct a WWW site which will incorporate the research they have done into a WWW resource or a user friendly tutorial. Students can create a unit or lesson plan which has the intended audience accessing archives and files all over the world; or a WWW site could be developed which sets up research for a paper in another area of study or for a thesis; or an on-line lesson plan that will satisfy the general requirements of Alberta Curriculum can be set up. It is crucial that students clearly define their audience and demonstrate why this resource will be beneficial to them. This is important because it helps them focus on a specific topic and audience and it also sets up the marking criteria. The most important consideration is that the final project must be incorporated into the students current work or area of specialty. A number of our studentsí final Web pages can be viewed at http://www.quasar.ualberta.ca/nethowto/articles/final.

Observations & Recommendations

There is no doubt in our minds that a very effective and efficient instruction can be delivered over the Web and that this course has moved in the right direction. Student learning, as evidenced by their comments and their assignments, was similar for students in the on-line version of the course and those in a F2F mode. We have learned a number of things while designing and delivering the course:

- the instructor should have taught the course previously in a F2F mode in order to design the course effectively
- use good adult learning principles (learner control of content, pacing & sequencing; alternative methods of learning; self-selected assignments)
- provide good student-student and student-instructor communications
- be prepared to deal with students who find the technology ìgets in the way of learningî
- make the course load as quickly as possible
- take advantage of other peopleís work (other resources on the Web)
- allow students to help each other and to discuss things among themselves without feeling the need to ìguideî all discussions, but provide fast feedback to students who are encountering problems.
- expect that the workload will be significantly higher than for a F2F course on similar material.
- be prepared to continually modify your course pages
- always remember, and remind your associates and students, that there are less cues in computer mediated communications than in F2F communications; what is said with a smile can sound harsh when printed
- enjoy the experience - it is different, but it is teaching!

References

[Montgomerie & Harapnuik 1996] Montgomerie, T.C. & Harapnuik, D. *EDPSY 497/597 - The Internet: Communication, Accessing & Providing Information.* Edmonton: University of Alberta. [http://www.quasar.ualberta.ca/nethowto].

[Levine 1995] Levine, A. (1995) *Writing HTML: a tutorial for creating WWW pages* Maricopa Center for Learning and Instruction. [http://www.mcli.dist.maricopa.edu/tut/lessons.html].

An Alternative to Browser-Based Instruction: Authorware Professional with Internet Extensions

T. Craig Montgomerie
Division of Technology in Education
3-104 Education North
University of Alberta
Edmonton, Alberta
Canada T6G 2G5
Phone: (403) 492-3667
Fax: (403) 492-3179
E-Mail: Craig.Montgomerie@ualberta.ca

Gary N. Duguay
Division of Technology in Education
3-104 Education North
University of Alberta
Edmonton, Alberta
Canada T6G 2G5
Phone: (403) 492-3667
Fax: (403) 492-3179
E-Mail: gduguay@gpu.srv.ualberta.ca

Background

There is a great deal of courseware being developed for delivery on the World Wide Web today using standard Web browsers. While Web browsers do many things extremely well, they are not specifically designed to provide direct instruction as are authoring systems such as Authorware Professional. Rather than try to force a Web browser to become an instructional system, we have added network capabilities to Authorware Professional to allow it to communicate over the Internet to load and update Authorware Professional modules, keep track of student restart data, student performance data and user license data. One of the side effects of using Authorware Professional rather than a web browser is that the source code is well protected since the user accesses object code only.

Computer Assisted Instruction (CAI) systems were originally developed and delivered on "mainframe systems" (1960s - 1980s). Students accessed courses via a terminal which was directly connected to the mainframe. This allowed the course author to make changes to a single version of the course and have that change take effect for all students. It also allowed a single, centralized registration and student data collection system. With the development of microcomputer-based CAI systems, each student can receive their own copy of the course, and can work through the course in complete isolation from the course author and system administrators. With the development of CD-ROM, students all over the world can receive a copy of a massive course of instruction, but the course is static - even if the authors find errors, they have to ship a new CD-ROM to all students to correct those errors - a relatively costly process and one which is not usually done expediently. Also, student data which could help improve the course can not be collected at a centralized location. The problem we must address is how we can provide the centralized services we used to have on mainframe systems in our current distributed computing environment.

Scope and Objectives

The purpose of this project is to develop a prototype file management system for Authorware

Professional courses which utilized the Internet as the medium for data transfer. The system automatically provides updates of course modules from the Server to the microcomputer and transfers student performance data from the microcomputer to the Server to help instructors improve CAI courses. Provisions have been made for on-site instructor intervention (to prevent module update, transfer of student information, etc.), and for operation of a license server. Specific objectives of the project include:

1. Identifying the kind of information which needs to be kept in an index to allow automatic checking and updating of course modules.
2. Identifying the kind of information which needs to be contained in a license database.
3. Identifying the kind of information which should be collected as students proceed through a CAI course (performance recordings).
4. Development of module index, performance recording and license databases on the central node.
5. Development of routines to automatically transfer update information over the Internet for both the remote microcomputer and the server.
6. Provide electronic communication between students and instructors.

Current Status

At this time we have identified the kinds of information which needs to be kept in a license database, that which is required to allow automatic updating of course modules and that which needs to be kept as student performance recordings. Databases with this information have been designed. A series of routines have been built which provide for the transfer of data between a stateless server (an IBM RS6000) and a computer assisted instruction program running under Authorware Professional on a Macintosh computer which has an IP address on the Internet. The routines have been built so that their operation is transparent to the user. The routines developed allow: course registration, logon and password for four levels (course owner, license holder, instructor, student) course licensing and version control, control of access to and updating of student/course data by instructors, maintenance of student restart records at the server, downloading of course segments, up/downloading of student data, and student/faculty communication. The routines have been built into the course *Understanding Statistics* which will be used to deliver a first year statistics course to graduate students at the University of Alberta in 1996-97.

How does it work?

A student or institution obtains the kernel or control module for the course either from a CD-ROM, through FTP or from a WWW site. CD-ROM is preferable, since a ìshrink-wrappedî product can be sold and original versions of each course module, graphics, etc. can be stored on the CD-ROM to minimize network traffic. Figure 1 is a simplified flowchart showing the interaction between the Student, the Client software and the Server once the Student launches the kernel.

Program Demonstration

A kernel for this program is available from http://grover.educ.ualberta.ca/~gduguay/demo/. This Macintosh demo will allow a user to register and download the files necessary to begin as a student in the Understanding Statistics course. This demo will remain operational until November 15, 1996.

Student/Client Server

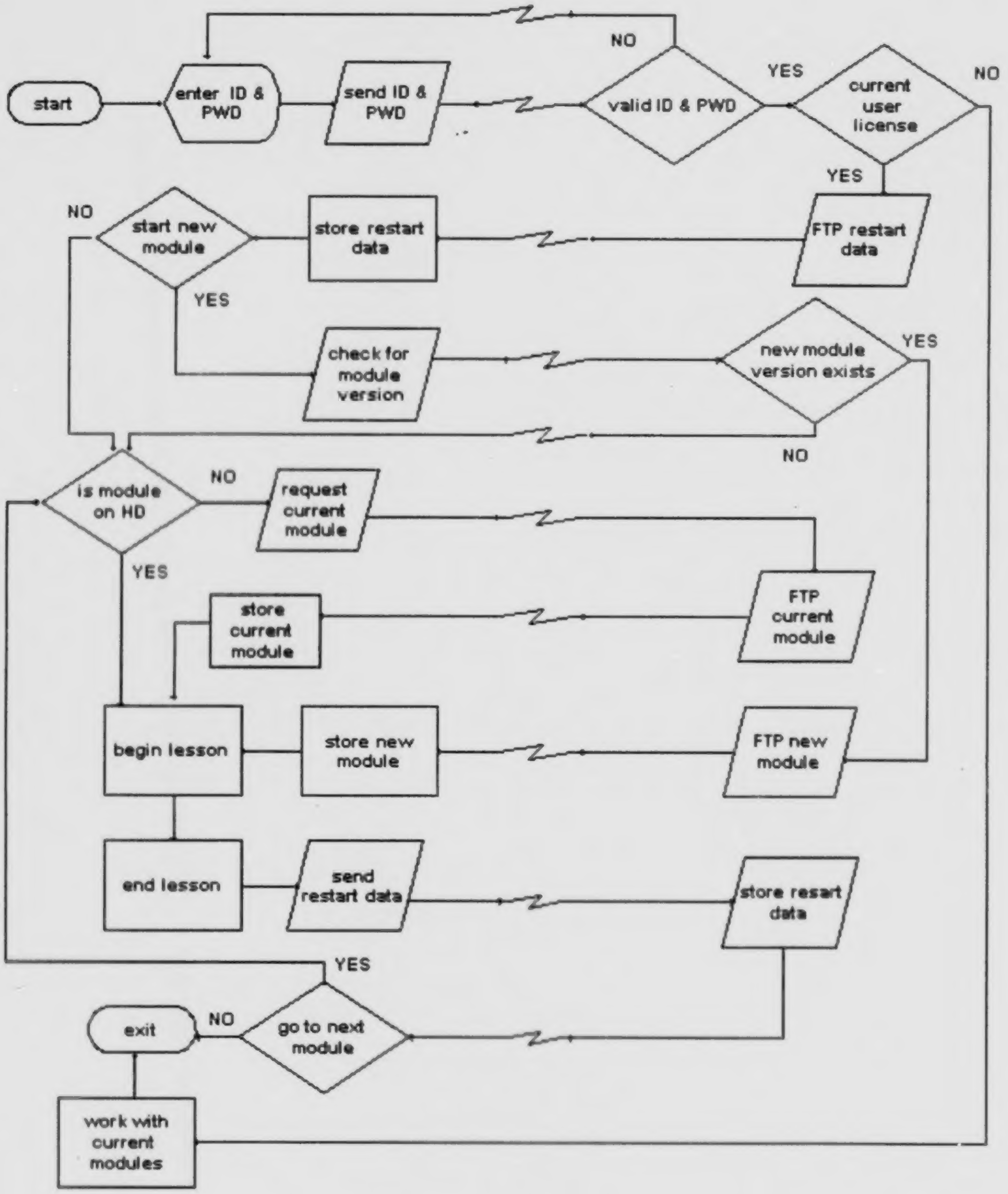

start
enter ID & PWD
send ID & PWD
valid ID & PWD
NO
YES
current user license
NO
YES
FTP restart data
store restart data
start new module
NO
YES
check for module version
new module version exists
YES
NO
is module on HD
NO
YES
request current module
FTP current module
store current module
begin lesson
store new module
FTP new module
end lesson
send restart data
store resart data
YES
go to next module
NO
exit
work with current modules

Figure 1: Student/Client/Server Communication

BEST COPY AVAILABLE

Using the World Wide Web To Creatively Support Science Education in Schools

Conway Mothobi
Telematics Research & Applications Centre, Department of Computing
Manchester Metropolitan University, Manchester M1 5GD United Kingdom
c.mothobi@doc.mmu.ac.uk
http://www.doc.mmu.ac.uk/trac/conway.html

Abstract: This paper reports on the construction of a framework that seeks to provide schools with a robust forum in which to effectively experiment with the World Wide Web and other Internet tools. It reports on the fragment of a pilot initiative which supports science education in schools. The initiative is primarily directed at providing schools with an interactive forum that permits cost effective learner-to-educator, educator-to-learner and peer-to-peer communications by providing;

- educators with a platform for effective use of Internet tools and providing learners with the opportunity to conduct science projects over the Internet using the World Wide Web;
- schools with access to such interactive resources as the Lawrence Berkeley Laboratories' Virtual Frog and the ACT Laboratory dens;
- schools with an interactive window into such real time science experiments as NASA's Live from Hubble Space Telescope (LHST) programme.

Introduction

The Faculty of Science and Engineering is committed to establishing mutually beneficial relationships with local schools and colleges. For the last two years the faculty has hosted a Science Day for schools during the UK National Science Week. Schools are encouraged to enter their projects in an exhibition with prizes donated by local industry. On Science Day schools visit the faculty and demonstrate their projects, and young people are encouraged to visit departments of interest to them to participate in some interesting activities.

Due to the university's own work programme and the schools' own schedules it is not possible to sustain such activities other than as special occasion events. In order to encourage on-going activities the faculty has now formed a free-to-join schools association. Schools can run their projects in collaboration with appropriate departments.

Initial support for these activities is being realised through the [Passport to Knowledge project] which facilitates access by schools and colleges to the faculty's departments via the Internet. These accesses enable the schools and colleges to collaborate effectively with the departments of their choice.

The Manchester Passport to Knowledge Project (PTK)

The project is based on the Passport to Knowledge concepts developed by NASA in their K-12 Internet Initiative (an Information Infrastructure Applications project). It is particularly influenced by the Livefrom Hubble Space Telescope (LHST) offerings for which a UK reflector was built in order to provide UK schools with a window into real-time science, technology and engineering.

The PTK project is essentially a framework for site constructions that fully integrate the Internet into the learning environment and schools provide a robust start point. The main communications tools are email, fax, WebChats, CUSeeMe and the Mbone. The key objectives of the project are to:

- provide learners (and educators) with a dynamic interactive environment that supports diverse and sometimes disparate learning styles;
- provide educators with a forum for effective curriculum discussions and development through such features as WebChats and discuss groups.
- ultimately provide learners and educators with fully interactive environments that support creativity.

3W Learner Activity

In PTK, learners are encouraged to pose three questions to themselves when using the web. These questions are based on the NASA 3Ws that were put forward during the LHST programme.

- *What* do I want to know?
- *Where* is this information to be found? Is there a resource I could easily check eg classmates, reference books etc
- *Why* do I want to know it and how will I use it?

In the LHST project it was stressed that the third question is probably the most important since learner reflection on the reasons for knowing is in itself a valuable and enriching experience.

Local Government Involvement

The Manchester Education Committee (Schools Board) has sponsored the cascading of web capabilities to schools by sponsoring a Young Persons Council (YPC). Although these activities are in the domain of "Democracy and Citizenship", their mediation over the web has been a major contribution in encouraging teachers to welcome the web into their working environments. This has been particularly so since the teachers have been able to pro-actively apply the web in developing new activities that are within the National Curriculum. The YPC is located at http://quest.doc.mmu.ac.uk/ypc/index.html

The YPC has enabled the PTK to: i) reach a much wider audience in the schools sector; ii) to experiment with the use of discuss groups within the UK school environment; and iii) to use CUSeeMe as a support feature for web activities.

Manchester Museum of Science and Industry

The PTK project has entered into a partnership with the local museum of Science and Industry. The main objective of the partnership is to present science as a fun subject, and the museum a fun place to be. University students on placement at the Museum produced some pilot activity packs for use by schools for visits to the Museum. These packs are available on the web and a school may download them and prepare a class for the visit. Availability of such material has been particularly useful in the primary school sector.

The PTK project and the Museum have now agreed to hold a schools slumber party on 15th-16th March 1997 as part of National Science Week 97. Activities being planned for the event include interactive use of the web with emphasis on WebChats and CuSeeMe. It is envisaged that pupils who are unable to attend will be able to participate via the Internet. It is envisaged that these activities will also involve the Houston Museum of Natural Science, whose expertise in CUSeeMe proved invaluable in our LHST activities.

PTK Sponsorship

Silicon Graphics Inc and Open Computers & Finance are now sponsoring our PTK activities with the provision of a server. This sponsorship marks a critical phase in the project as we are now in a position to develop an interactive education site (edusite). Such developments include the involvement of final university learners in virtual worlds projects with a view to developing appropriate 3D environments that are suitable for the dissemination of science - in particular it is envisaged that the work will also increase our activities in collaborative engineering projects with the private sector

Summary

In summary, the goals of the PTK are to establish an interactive edusite for the support of science education.

Bibliography

Mothobi, C and Strom, J (1995) Using the World Wide Web as an Education Support Tool. Proceedings of the 3rd Annual Conference of the Computers in Teaching Initiative (CTI). CTI Jodarnstown.

Strom, J and Mothobi, C (1995) A Broadband Infrastructure for the Educational Community in Greater Manchester. Proceedings of the 4th International Conference on Broadband Islands. Vesticom Dublin.

Acknowledgements

Thanks and appreciation are due to: Manchester City Council, Silicon Graphics Inc, Open Computers & Finance and the NASA K-12 Internet Initiative.

OODB Support for WWW Applications: Disclosing the internal structure of Hyperdocuments

J.T. de Munk, A.T.M. Aerts and P.M.E. De Bra
Department of Mathematics and Computing Science
Eindhoven University of Technology
PO Box 513, 5600 MB Eindhoven
The Netherlands
{munk,wsinatma,debra}@win.tue.nl

1. Introduction

Most World Wide Web (WWW or Web) servers use the operating system's native file system for the storage of the HTML documents and their embedded images. This access mechanism works fine for direct access, but is ill suited for finding documents containing certain information. Even just finding all documents that exist on a server is difficult because of the complexity of the hypertext link structure. The structure may not be completely connected, meaning that some documents on a server may not even be reachable by following links from other documents.

Several attempts have been made to build additional structures, that provide search facilities for the information stored on a WWW-server or a cluster of servers. Glimpse [Manber & Wu 94] provides indexing at the server level. Harvest [Schwartz et al. 94] extends Glimpse to offer retrieval over a set of servers.

The existing index databases ignore most or all of the internal structure of the documents. Asking for information that appears in a "header" of certain levels, e.g. levels 1, 2 and 3, is not possible. Finding information in an <address> field is impossible as well.

The source of the problem is the lack of a sound access-mechanism for the information one is interested in. The flat file system approach taken by Web servers makes it easy to access a whole document, given its address, but makes associative retrieval difficult. An "inverted" access mechanism is needed, providing access to documents or parts of documents, given a description of their contents, their internal structure, or the link structure of their environment.

Instead of adding an index-database onto a file system based Web server, we propose a server based on an object-oriented database, delivering the documents and the answers to search requests from the same information source. Documents are stored as objects of which the internal structure represents the HTML structure, thus enabling querying for structural elements like headers, hypertext links, quotes, addresses, etc.

The most influential similar approach to enhancing the WWW is the Hyper-G project [Andrews et al. 95], which has a richer structure than WWW, but can reduce documents to HTML in order to serve them to WWW browsers. We take a different approach by keeping the WWW architecture, to the point where we integrate an existing WWW server with an object oriented database system.

2. Requirements and Properties

When we started the development of the new server architecture, the following requirements guided the process, and resulted in the corresponding properties in the prototype system:

- The new server had to be built using freely available technology wherever possible. This requirement resulted in the following choices:

  - Instead of building a completely new WWW server, the code for the CERN server was used. Two versions of the new server have been built: one that leaves the CERN server code completely intact, and one that uses a small modification to eliminate the need for an additional CGI-script (which slows things down).
  - The Ode database system was chosen [Agrawal & Gehani 89], [Gehani 91], because it was and is freely available to universities (after signing a license agreement with AT&T).
  - The HTML parser needed to be at least as forgiving to syntax errors in HTML documents as most browsers. For this reason we have developed our own HTML parser, instead of reusing an existing parser like SGMLS or SP [Clark 95].
- The WWW server of our department would be the first test case. Since we have many people serving documents off that server, the transition from the file system based server to the Ode based server had to be very smooth. We have therefore developed a "two headed" server, which can serve documents from the file system and from the database simultaneously and transparently. The user cannot deduce from the address (URL) of a document, or from its appearance, whether it is served from the file system or the database. In fact, with scripts like the common *imagemap* program, it is possible to use the existing (file system based) script to associate the coordinates of a mouse click to a document, and to subsequently serve the document from the database.
- The server needed to parse HTML documents to store them internally as structured objects, but needed to be able to reproduce the documents exactly as they were input, including details that are insignificant to HTML, like the use of upper- or lowercase in HTML commands, the order of attributes to commands, additional white space and the position of newlines. By realizing this goal through the object structure used in the database it becomes easy for users working on documents together to retrieve a document from the server and compare it (using *diff* or a similar program) to their local version in order to find modifications made by their coauthors.

3. Future Work

We have built a WWW server which stores the internal structure of documents in the object oriented database Ode. A primitive search facility has been added to this server. The internal index structures to make searching efficient will be added in the near future.

We also wish to integrate our document repository system DReSS [BA95] with this server. By doing so the basic unit for locking can be reduced from a whole document to the smallest structural HTML elements, enabling more concurrent authoring than with the current version of DReSS.

4. References

[Agrawal & Gehani 89] Agrawal, R., Gehani, N.H. (1989). Ode (Object Database and Environment): The language and the datamodel. Proc. ACM SIGMOD Int. Conf. on Management of Data, pp. 36-45.

[Andrews et al. 95] Andrews, K., Kappe, F., Maurer, H. (1995). Serving Information to the Web with Hyper-G. Third International WWW-Conference, Darmstadt. (http://www.igd.fhg.de/www/www95/papers/105/hgw3.html)

[BA95] De Bra, P.M.E., Aerts, A.T.M. (1995). Multi-User Publishing on the Web: DReSS, A Document Repository Service Station. Proc. NLUUG Autumn Conference, Publishing on the World Wide Web, Ede, 13-26. (http://www.win.tue.nl/win/cs/is/debra/dress/dress-paper.html)

[Clark 95] Clark, J. (1995). SP, An SGML System Conforming to International Standard ISO 8879 -- Standard Generalized Markup Language. (http://www.jclark.com/sp/index.htm)

[Gehani 91] Gehani, N.H. (1991). The Ode Object-Oriented Database Management System: An Overview. Advances in object-oriented database systems, 355-387.

[Manber & Wu 94] Manber, U., Wu, S. (1994). Glimpse: A Tool to Search Through Entire File Systems. Proceedings of the USENIX Winter Conference. (http://glimpse.cs.arizona.edu:1994/)

[Schwartz et al. 94] Schwartz, M.F., Bowman, C.M., Danzig, P.B., Hardy D.R., Manber, U. (1994). The Harvest Information Discovery and Access System, Proceedings Second WWW Conference, Chicago, 763-772. (http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/schwartz.harvest/schwartz.harvest.html)

An Internet Information Service for the Domain of Personnel and Development

Sheila P. Naujoks
School of Cognitive and Computing Sciences, University of Sussex, UK,
email: sheilan@cogs.susx.ac.uk

Laurie Griffiths
Personnel and Development Communications Ltd., UK,
email: 100105.1355@Compuserve.com

Theodoros N. Arvanitis
School of Cognitive and Computing Sciences, University of Sussex, UK,
email: theoa@cogs.susx.ac.uk

The Domain

Personnel and development are inherent concerns of any organisation. Every company, no matter how small, has to recruit, train, assess and pay their workforce. This is where the demand for information, products and services for this sector comes from.

At present, personnel managers need to access many different sources of information - directories, publications, specialist consultants and other information services - in order to fulfil all their varied tasks. This is very time consuming and often costly and it therefore would be desirable if there was one source of information which would provide everything.

The Internet and Personnel and Development

The Internet and particularly the World Wide Web could be this single source of information, catering for all the needs of personnel and development. This is because it is widely accessible - all one needs is a computer and modem; most companies have the former and an ever increasing number is getting an Internet connection. There is no need anymore to store vast amounts of publications and no time is wasted to find relevant information amongst them.

In contrast to static information contained in any other publications - be it electronic or on paper - WWW pages can be "up-to-the-minute" and therefore the user will always be presented with correct information and the latest updates. This is important in personnel and development areas such as employment law and useful for announcements regarding changes of place or date of conferences, for example.

However, as the WWW is still in its infancy, its structure and organisation leaves a lot to be desired. Users that are not going to a specific URL but instead browse and search the WWW to find information about a certain topic, tend to find that they have to move through a vast amount of 'junk' before they get to a useful web page. Given that many prospective users of our information service may not have had much exposure to computers or even are computerphobic, we anticipate that they will experience problems finding relevant information. This might discourage them from using the WWW.

In order to avoid such problems, the information service has to be user-friendly and efficient: Our system presents data in a structured way, based on the users' cognitive model. An efficient search facility assists the user in navigating through the web-site. Human-Computer Interaction techniques have been employed to ease-of-use. This lead to the construction of graphical user interfaces. These measures will eliminate the potential for frustration the novice user might experience.

All relevant information and links to other useful sites have to be included in order to make our system stand out from already existing ones and to avoid that the user has to search the WWW for missing information. It is also important that our information service appears at the right places, with a meaningful abstract, for people using a WWW search engine.

The System and the Proposed Solution

The web-site will ultimately contain all the information our users need, as well as links to related web pages. In order to facilitate navigation within the web-site a search engine, based on intelligent agent technology, assists the user in finding relevant pages.

For example, Alan Miller, Personnel Manager for Allied Breweries, specifies his preferences to be recruitment, training and testing & assessment upon registering with the service. When he does a search on computer software the same suggestions will come up as for other users, but the order will be different: software related to recruitment, training and testing & assessment will be put first.

According to our user study a significant number of prospective users prefer other types of searches to keyword search and we will probably get many people 'browsing' rather than searching for something in particular. We therefore asked our sample to propose structures for topics of their interest in order to be able to build our site in accordance with the mental model of our average user.

Implementation

When the user first enters the site, she has the opportunity to register. This is done to benefit from the personalised agent. If she chooses to do so a graphical user interface comes up, prompting her to enter her details and preferences. The interface is programmed in Java and looks similar to many HTML forms. A new record is created and the information entered by the user is stored in the engine's knowledge base. The user can decide on a name under which the agent will 'recognise' her. In order to use the search engine, the user has to enter her name, which allows the search engine access to her details and preferences. Then she can enter up to three keywords of the information she is looking for into the keyword field of the search engine interface.

A database stores information about each page: its URL, and up to 8 keywords that describe the contents of the page. The keyword(s) entered by the user are first compared to a list of synonyms of all keywords in the database and replaced with the term chosen for the database. For example, the user might have entered the words jobs and education. The corresponding terms in the database are recruitment and training respectively. Each term, however, has a list of synonyms (e.g. recruitment: jobs, employment, vacancies, …) and if the user's keyword is found therein, it is replaced by the database word (e.g. jobs is replaced with recruitment) before the search engine searches the page database.

The user's keywords are compared to the keywords of the pages and three lists are created: The first with all those pages which have three of the user's keywords among their keywords, the second with those that have two keywords in common and a third list with pages that only have one match. The lists are then sorted according to known user details and preferences. For example, someone looking for 'employment law' and 'solicitor' will get all pages with that description presented before the pages that have only one of those keywords linked to them. Furthermore, those pages with solicitors in the same city or county will be put first. Equally, when a user who specified 'training' as one of their main interests runs a search on 'computers' and 'software' she will get pages related to training software put before pages with lists of dealers or software developers. Once the lists are sorted they are displayed to the user: the one with three matches first, followed by the one with two matches and the one with one match last. The user can then go to the suggested pages by clicking on their highlighted URL.

Future work

The ultimate aim is to enable our software agent to learn, i.e. to store information regarding how long a user spends on certain pages upon which a user model can be built. By 'watching' the user, the agent will not only be able to learn the user's interests and therefore present the most useful pages first, but can also draw conclusions from existing user models in order to better present search results to new users.

A Virtual Classroom in the Web

Otalora, German (gotalora@campus.mty.itesm.mx)
Herrera, Ramses (jherrer@campus.ruv.itesm.mx)
Monterrey Institute of Technology
Monterrey. Mexico

Introducation

For the past six years Monterrey Institute of Technology's Virtual University has been developing a variety of virtual and distance learning experiences and has experimented several ways to deliver education to the 26 campus located all over Mexico and to some sites abroad. One of these experiences is the capstone course for undergraduates majoring in Business and the Strategic Management Course for MBA students.

This paper will discuss only the undergraduate course. During the current term, students located in six different cities can get together in a virtual classroom constructed from a home page in the web.

The only paper materials for this course are the textbook (Hill & Jones. Strategic Management) and the student's manual for the simulation (Cotter & Fritzsche. Business Policy Game). All other course materials, including syllabus, program, assignments, course policies and rules, instructions, etc., are provided by a WWW server and a home page.

The Home Page

The whole learning experience has been designed following the analogy of a study trip. Therefore, the home page shows an airline counter through which students document themselves and get their boarding passes. This registration step allows the course administrators to have control of those entering to the links to the pages where the course materials are placed. Once they have registered, after a welcome message, they get access to the links in the home page. The first link leads to the calendar and schedules; the second, to the syllabus (fly plan); the third, to the pages containing safety rules for the trip. Membership miles are computed for each of the 117 students enrolled in the course.

A good performance deserves an upgrade to first class. There is a link to the first class passengers roster. Finally, an authorized personnel door takes you to the pages reserved for the seven professors involved in this course.

The Learning Experiences

The capstone course for Business majors has these learning experiences:

* Virtual groups, simulation, cases, discussion group, and face to face and distance classes.

* Virtual groups are small teams of three to four students, each from a different campus, communicating by e-mail and exchanging documents by FTP. All of the groups have the same assignment. The group's roster, assignment's questions, a link to the FTP site, policies and procedures are all found in the web. There is also a link to the company's pages which they are analyzing.

The Business simulation is run by our Simulation Master in a campus located at the Mexico City's metropolitan area. Decisions are send by FTP and results are obtained also by FTP. These results are transfered to the game's diskette from which students can run graphs, market and financial analysis.

One case is assigned for group discussion every week. The case questions and links to the company's pages are provided in an specific page.

We have created a discussion group to which all the students and professors can send their comments. A lead question is presented to the group, and contributions are evaluated by all the participating professors.

We have classes twice a week. On Tuesdays the class is face to face. It is dedicated to a brief lecture and case discussion. On Thursdays we have a general satellite meeting conducted by the professor responsible of the whole experience. During these satellite sessions we have a Newscast, a short visit to a host campus and some of the main companies in that location, a review of the theory, a video supporting the case discussed on Tuesday, a summary of the case discussion and a brief presentation of the theory to be covered the following Tuesday .

Every other Tuesday we have a video link with the simulation Master for comments and doubts about the game. During the satellite session we have written and voice mail for doubts and comments.

An e-mail address is open to all the members of the group for communication during the semester.

Software and Supporting Infrastructure

The course is supported by two kinds of communication:

* *On line*. This refers to all the technologies that are used during the class session: telephone, fax, live satellite transmision and a real time remote interaction system which the students and teacher use to make and answer questions about the class.

* *Off line*. This refers to all the technologies that are used out of class: e-mail, Web pages, Net News, FTP servers, distribution lists using majordomo and a business simulation program.

This course stands for the efforts that the Monterrey Institute of Technology is doing to change the actual educational paradigm by developing a distance educational program by using avant garde technology in all of its courses.

Put Your Experience on The Web: A Tool for Creating and Browsing in Case Libraries

VALERY A. PETRUSHIN, JANET L. KOLODNER
EduTech Institute,
Georgia Institute of Technology
Atlanta, USA
e-mail: {petr, jlk}@cc.gatech.edu

Introduction

This paper describes a class of AI software called Case-Based Aiding Systems (CBAS) and a tool for creating and using such systems in both off-line and on-line modes. CBAS is intended for improving human decision making using Case-based Reasoning approach [Kolodner, 1993]. According to [Kolodner, 1991], a CBAS would provide cases to human problem solvers at appropriate times to help them with such tasks as coming up with solutions, adapting old solutions, and warning of potential problems. A CBAS looks like a smart database. It retrieves cases based on user requests. But it is more than a database because it uses more sophisticated searching methods for finding partially matching cases.

A specific class of CBAS called Case-Based Design Aid (CBDA) has been developed over the past 6 years at the Georgia Institute of Technology. A CBDA is ìa tool to help a designer produce designs more successfully and more quicklyî [Domeshek, 1994]. The CBDA serves as common memory for designers. It allows the designer to retrieve relevant cases and to index and store a new design creating a synergetic power of human-computer interaction. It does not automate the process of design but serves as a source of inspiration and creativity for designers. Owing to efforts of many researchers, a shell Design-MUSE for building CBDA has been implemented [Domeshek, 1994]) and several CBDAs for different problem domains have been created: buildings design - Archie and Archie-II [Domeshek & Kolodner, 1992], planning lessons - SciEd [Kolodner, 1991a], aircraft subsystem design - MIDAS [Domeshek, Herndon, Bennett, & Kolodner, 1994], sustainable technology - SUSIE. The above projects showed the feasibility and range of applicability of CBDAís and taught many lessons about how cases needed to be represented and what the user's interface should look like.

The CBAS can be used to solve a wide range of problems in business and education. The salient feature of a CBAS is its extendibility. A CBAS may serve as a log or dairy of your professional or social growth from novice to expert. It is your extra mind. It can also serve as a common shared memory for a group of experts or a corporation. In education a CBAS can serve as a source of knowledge when students browse it and as a cognitive tool when students create their own CBASs.

ExtraMind: a Shell for building Case-Based Aiding Systems

ExtraMind inherits many features of the Design-MUSEís case library structure and is designed to be more generally-applicable and run on both platforms (Macintosh and IBM PC) and on the WWW. Like Design-MUSE, it can be used to build CBASs. Such a system consists of a library of cases, retrieval engine, and userís interface. The case library contains of a set of cases. A *case* may represent an artifact or an event. Each case may have parts (subcases) and/or views. A *part* can also have parts and *views*. There are no limits on the number of parts or nesting level. A part may belong to several superparts or cases. Cases, parts and views form a (partial) hierarchy of a CBAís memory. Each case, part, or view can have links to stories and sources. A *story* is an evaluated description of particular artifact or event which relates to the case(s) and illustrates problems and responses to the problems (decisions). Each story (and especially negative ones) should teach a lesson to the user. A story has links to the relevant cases (parts and/or views), problems and responses. A *problem* presents the descriptions of some general difficulties which relate to the stories. A problem is tied to the relevant responses, stories and cases (parts and/or views). A *response* presents a general strategy which has been applied to solve a problem. It also has description and links which tie it to relevant problems, stories and cases. A *source* is a reference to a source of information used in a case (part or view), story, problem or response. It may be a bibliographic citation or a name of expert. Each of the above mentioned objects has its own index which labels the object and can be used for retrieval.

There are two basic approaches to finding information in a CBAS: searching and browsing. In searching, a user specifies what he/she is interested in finding by specifying some physical or functional parts of a system (for example, work area and lighting) and an issue or set of issues (for example, natural light). In response, the system finds cases or stories, that address the specified issues (in this case, a story about vaulted ceiling and clerestory windows in the Roswell library). To make a story fully understandable, the system presents several other things as well: (1) the generic problem the story is about (bringing in natural light), (2) the type of solution it implements (for example, using atriums and indirect light shafts), and the larger case the story is part of (the Roswell library design). In browsing, a user can follow links within the case library to look at a series of stories, problems, solutions, and cases. Linkages within the system allow easy traversal from stories to other stories that illustrate the same problem or solution; from stories to other stories from the same larger case; and from problem to their alternative solutions. Combining the searching and browsing modes the user can explore the cases, stories, problems and responses which are relevant to his/her goals to acquire past experiences, learn the lessons taught, and adapt the past decisions to his/her new problem.

Implementation

Having started the development of a tool for creating and browsing CBASs we had the strong feeling that a CBAS can be considered to some extend as an aggregate of a database and a hypermedia system. In fact, a CBAS, on the one hand, is a special kind of database based on the cognitive theories of memory, but, on the other hand, it is similar to a hypermedia system because it heavily uses associative links. Besides that we tried to meet some other requirements such as: (1) the tool should work in both off-line and on-line (WWW) modes; (2) the off-line mode should be supported on both platforms: Macintosh and IBM PC, which are popular in academic environments; (3) the user interface should look similar in both the off-line and on-line modes.

The above intuition and requirements led us to use a multi-platform database management system (DBMS) with a flexible scripting language for implementing the off-line version of the tool, and to use HTML and Common Gate Interface (CGI) programs for implementing the WWW version. This decision allowed us to meet the requirements but also brought many other advantages: first, we could borrow effective searching techniques implemented in modern DBMSs; second, we did not have to reimplement security features for the CBAS tool; third, the scripting language could allowed for the easy creation of user interface. FileMaker Pro has been chosen as a multi-platform DBMS to implement the first off-line version of the CBAS tool. The additional advantage of this choose is the possibility to share case libraries on a local network among users on different platforms. We have chosen Macintosh as platform for publishing CBAS on the WWW. Currently we have implemented the CBAS browser for the WWW. The next step is to implement an authoring tool for creating CBAS using WWW clients. This will allow for a distributed team of experts to work on common projects.

There are several other technical aspects: (1) to implement more sophisticated search techniques, which gain more from the hierarchical index vocabulary structure; (2) to implement some adaptation and evaluation techniques which will help the user to solve a new problem; (3) to improve the user interface using the abilities of advanced WWW clients; (4) to add scaffolding that will help both the user to solve problems and the experts to create or modify CBASs.

Additional information about ExtraMind and a sample case library SUSIE which was created for learning sustainable technology can be found at http://mime1.marc.gatech.edu/WebCBRPage.html.

Acknowledgments

This work was supported by the EduTech Institute through a grant from Woodruff Foundation.

References

Domeshek, E.A. 1994. *The Design-MUSE v2.0 Useris Manual*.

Domeshek, E.A.; Herndon M.F.; Bennett A.W. & Kolodner, J.L. 1994. A Case-Based Design Aid for Conceptual Design of Aircraft Subsystems. In *Proceedings of the Tenth Conference on AI for Application*, IEEE Computer Society Press: Washington, 63-69.

Domeshek, E.A. & Kolodner, J.L. 1992. A case-based design aid for architecture. In J. Gero (ed.) *Artificial Intelligence in Design 1992*. Boston: Kluwer Publisher.

Kolodner, J.L. 1991. Improving human decision making through case-based decision aiding. *AI Magazine*, 12(2), 52-68.

Kolodner, J.L. 1991a. Helping teachers teach science better: Case-based decision aiding for science education. In L. Birnbaum (ed.) *Proceedings of the International Conference on the Learning Sciences*. Charlottesville, VA: Association for the Advancement of Computing in Education.

Kolodner, J.L. 1993. *Case-Based Reasoning*. Morgan Kaufmann Publishers, Inc.: San Mateo, CA.

The CyberQuest Project: A Pilot Project Connecting K-16 Educational Institutions to Internet via High Speed Data Access

Dr. Marvin L. Peyton, Department of Educational Leadership, Middle Tennessee State University, USA.

mpeyton@frank.mtsu.edu

Overview:

After several months of research and planning the CyberQuest Project was initiated in June of 1995. This project utilized high speed connectivity 128Kbps band width and not dial-up PPP access at 56Kbps band width.

The CyberQuest Project Model is currently being used a model by the State of Tennessee as it provides a means for total connectivity for all 1554 K-12 public school buildings (via the state's ConnecTEN Project) to a county Point of Presence (PoP) router and/or modem. The CyberQuest Project is providing a functioning networking model that currently provides its participants access to the Internet.

The CyberQuest Project goals.

- establish a cost effective network of services that are not currently available to local educational entities at a reasonable cost
- provide local access to the users
- provide a variable bandwidth that can grow to meet the end-users needs
- establish an electronic network that can and will exist into the future
- provide a fully functioning model
- enable end-users a method to blend the local resources of entities such as, libraries, hospitals, local government institutions, K-12 schools, etc.

Site Selection:

Selection of participating schools was determined first and foremost by their proximity to the central "enterprise" hub. Each participating institution is required to provide as a minimum the following hardware and /or personnel:

- One PC 486/33 500HD or greater or One comparable Macintosh
- Dedicated server location (secure)
- Individual we shall call the Point Person
- Two Phone lines (1 voice, 1 ISDN [CyberQuest will pay for this service])
- To facilitate network connectivity (internally) a service distribution hub for their teachers and students access.

The CyberQuest Project provided to each participating site:

- one ISDN router
- configure the site server
- install software for the LAN for the site
- maintain the site router for one year
- train the point person in the hardware and software specifics
- provide training opportunities for the site personnel
- provide one year ISDN service (pay the monthly connection fee)

- provide Internet Domain Name Service
- provide IP address
- provide email POP3 instructions and assistance for setup
- provide instruction and software for Internet services, Archie, FTP, etc
- provide news service software
- establish Internet connectivity via central router hub

The project provided the following training for participants:

- ensure skills and knowledge to use the equipment
- in use of resources available on the Internet
- methods to allow the participants to incorporate Internet into their curricula
- a mechanism for collaboration of success(s)/failure(s) in the design, implementation, and utilization of the Internet and specifically the WWW as part of their educational experience.

Project Implementation:

The project was initiated at Freedom Middle School in Franklin, Tennessee. The connection and LAN worked with virtually no failures. All participants are currently on-line and using their connectivity via multiple platforms and most have internal LANS.

Outcomes:

The project is operational and fully functioning at this date. The installations are finished and the training are ongoing.

Major Project Obstacles and Some Solutions:

The most common obstacle to viable connectivity to the Internet was that one presented by unprepared, untrained or uninformed teachers and administrators (not by technology or even funds for technology) when the concept of internet connectivity was presented. Fortunately, this obstacle was overcome in all most every instance.

Tennessee schools have very few individuals with a minimum degree of understanding of computer networking and less of an understanding of the hardware and software requirements to connect a LAN to the Internet. To circumvent this problem Tennessee has created a special technical position for each school district. This group of Technical Coordinators (TC) at least one TC for each Local Education Agency (140 LEA's) and over 240 TC's are those responsible for all the microcomputer technology in their individual district. This added responsibility for connectivity is not always welcome, because the TC's are in many instances over-worked and under-funded.

When the technological obstacles were met and the point person was comfortable with the "new" technology the initial use of the new resource was very limited.

Conclusion

Every school building is very different in it's construction. This has created a need for a model for connectivity but has in effect eliminated the prospect for a LAN model that truly is reproducible. Building access to conduct a pre-visit inspection or later installation was a major problem. Each building brings with it a special set of construction and wiring circumstances that must be dealt with by a specific plan for that edifice.

Recommendations from the Study/Pilot Project

A willingness by primary decision makers in the K-16 academic arena to support such a project financially and more importantly academically is the key to the success of the implementation of the

Internet and it's world wide resources into the K-16 curriculum.

Incorporating World Wide Web technology into a multimedia learning environment

Richard Picking
School of Computing
Staffordshire University
Stafford ST18 0DG
England

tel: 0044 1785 353444
e-mail: r.picking@soc.staffs.ac.uk

Introduction

This short paper describes a hypermedia environment to support essay and dissertation development for music students. The hypermedia system has been named HECTOR, an acronym for *Hypermedia from Essay Conception TO Realisation.* HECTOR is a composite system which incorporates a number of novel artefacts as well as some off-the-shelf products. It provides an integrating environment for these artefacts, and a structure which reflects a model of essay and dissertation development proposed in earlier work [Picking 1996]. The system is still under development, but executable prototypes have been tested at the City of Leeds College of Music (CLCM) in the UK.

The rationale for HECTOR

The study of the writing process has been of interest to cognitive psychologists for some time. Many models have been proposed, the best known of which is perhaps that of Hayes and Flower [1980]. A number of specialised computer-based writing and argumentation environments have recently been developed, for example gIBIS [Conklin & Begeman 1988], WordProf [Ferraris et al. 1990], SEPIA [Streitz et al. 1992], MediaText [Hay et al. 1994] and the CLASS Project [Gunn et al. 1995]. Such tools tend to be focused particularly at the cognitive level of writing, and in this respect they are based very much on cognitive models of the writing process [Buckingham Shum & Hammond 1994]. They do not consider the influence that other artefacts may have upon writing strategies, for example the availability of research material and alternative media requirements. Certainly, in education, reading and writing go hand in hand, as one must read to learn before demonstrating any newly acquired knowledge.

Earlier work by Picking [1996] in the field of music education has resulted in the proposal of a classification model of essay and dissertation development. The model was developed at CLCM through semi-structured interviews, which utilised the technique of repertory grid analysis [Kelly 1955; Bannister & Mair 1968]. The interviews enabled students to provide their strategies, perceptions and preferences of essay and dissertation development. The analysis of the data collected from this study resulted in a three-dimensional typology, broadly covering activities related to research, planning and writing, referred to as type I, type II and type III respectively. Consequently, HECTOR's artefacts support research (type I), planning (type II) and writing (type III) activities.

HECTOR enables users to structure and annotate multimedia from a number of information artefacts, such as the World Wide Web (WWW), CD-ROM and music compact disc servers. Users can seamlessly perform type I, type II and type III activities within HECTOR's graphical environment. The artefacts supported by the current prototype of HECTOR are summarised below. Figure 1 displays a typical screen from a session with HECTOR.

Type I Type II Type III

Netscape browser Overview Word processor

CD player Bibliography

Music reader Concept mapper

Jazz encyclopaedia

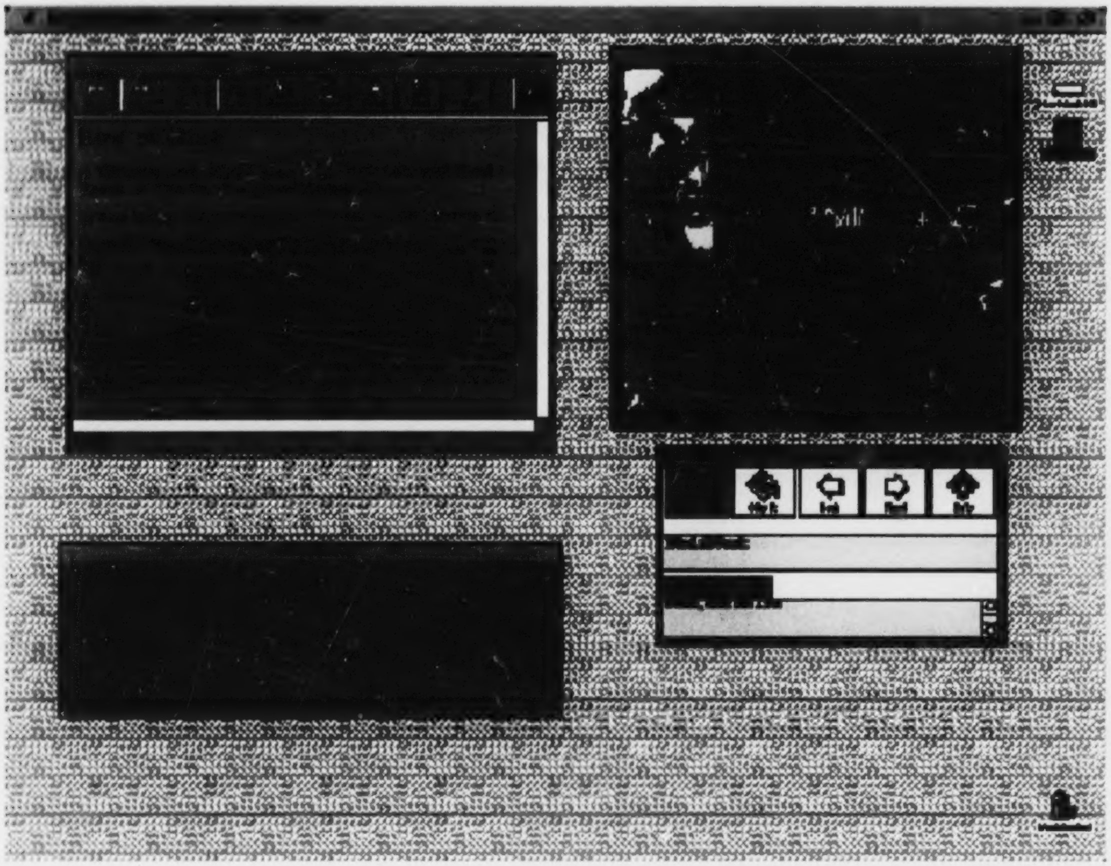


Figure 1: Typical screen from HECTOR

References

[Bannister & Mair 1968] Bannister, D. and Mair, D. (1968). *The Evaluation of Personal Constructs*, London: Academic Press.

[Buckingham Shum & Hammond 1994] Buckingham Shum, S. and Hammond, N. (1994). Argumentation-based design rationale: what use at what cost?, *International Journal of Human-Computer Studies*, 40, 603-652.

[Conklin & Begeman 1988] Conklin, J. and Begeman, M.L. (1988). gIBIS: a hypertext tool for exploratory policy discussion, *ACM Transactions on Office Information Systems*, 6(4), 303-331.

[Ferraris et al. 1990] Ferraris, M., Caviglia, F. and Degl'Innocenti, R. (1990). WordProf: a writing environment on computer, *Educational Training and Technology International*, 27(1), 33-42.

[Gunn et al. 1995] Gunn, C., Granum, G. and McAleese, R. (1995). The student organizer: a hypermedia

BEST COPY AVAILABLE

tool for effective learning, *Proc. Conference Hypermedia at Work*, Canterbury, UK, 10-12 January, 82-88.

[Hayes & Flower 1980] Hayes, J.R. and Flower, L.S. (1980). Identifying the organisation of writing processes, In L.W. Gregg and E.R. Steinbert (Eds) *Cognitive Processes in Writing*, Hillsdale, NJ: Erlbaum, 3-30.

[Kelly 1955] Kelly, G. (1955). *The Psychology of Personal Constructs*, 2 Vols. New York: Norton.

[Picking 1996] Picking (1996). *The potential for hypermedia to support educational practice in a college of music*, Unpublished PhD thesis, Loughborough University of Technology, UK.

[Streitz et al. 1992] Streitz, N., Haake, J., Hannemann, J., Lemke, J., Schueler, W., Schuett, H. and Thuering, M. (1992). SEPIA: a cooperative hypermedia authoring environment, *Proc. ECHT'92 4th ACM Conference on Hypertext,* Milan, November 30 - December 4, 11-22.

Non-Invasive Improvements to Hyperlinking with HTML

Maria G. Pimentel
Computer Science Dept.
USP/São Carlos, Brazil
mgp@icmsc.sc.usp.br

John Buford
Computer Science Dept.
U.Mass Lowell, USA
buford@cs.uml.edu

1 Introduction

HTML inherits from earlier hypertext both the link-node model for associating external documents, and the presentation and interaction behaviour provided by HTML document browsers. HTML document navigation is by means of basic operations defined for hypertext: free choice of link selection complemented by backtracking, bookmarks and history mechanisms, to name a few.

The simplicity of the HTML model, however, limits the link structure that can be associated with the underlying document collection. To illustrate the problem, this paper presents some linking structures well known from the previous hypertext systems and which are not supported in HTML. We then refer to experimental work in which these constructs are supported by Java applets without modifying HTML.

Due to space constraints, we omit a detailed review of the limitations of HTML links in comparison with widely discussed second generation hypertext systems. A number of other researchers have discussed improvements to the link model of the web, including [Maurer 94] [Ladd 95] and [Carr 96]. Many proposals require either changes to the current web architecture or extensions to the server such as link databases. The current discussion focuses on non-invasive improvements; in the long term we believe that more extensive improvements to the existing HTML hyperlink paradigm will be desirable.

2 Alternative Hyperlinking

Multi-end Link. One interesting construct used in Intermedia [Meyrowitz 86] was the link with multiple destinations, also known as fat link. The author could specify a source anchor with several destinations; at the time of selection, the user is presented with a pop-up menu listing the possibilities and can choose between them. Another approach to deal with fat links has been programmed by users of NoteCards: when the link is selected, one window is opened for each of the destinations [Halasz 1988].

Pop-up Link. A temporary pop-up link is such that, when its source anchor is activated, the corresponding destination anchor is displayed on top of the original anchor (and its window) for a limited period of time. This type of link was available in systems such as Apple's HyperCard and OWL's Guide. The temporary Pop-up link is as part of the HTML+ specification; however, current browsers do not support it.

Previewing the Destination of a Link. The previewing of the contents of the destination anchor was originally used in HyperTies [Shneiderman 87] as an alternative to the user not to move from the context containing the source anchor. When link traversal is initiated, a resumé of the destination anchor is displayed. Afterward, the user can choose whether to finish the selection. The resumé proposal has been expanded and generalized to allow cyclic access to information about the associated anchors [Pimentel 94].

3. Allowing alternative hyperlinking in HTML documents

One way of allowing HTML documents to achieve the alternative linking without extending the language itself is to use Java applets to implement the linking specification. As a result, the language is kept simple and the extensions are achieved by the use of simple applet activation. We next comment how such specification can be done for the alternative link types discussed (more details in [Pimentel & Buford 96]).

Multi-end Link. The Intermedia pop-up menu approach could be used to allow for multi-end links in WWW pages. In this case, a Java applet is implemented that receives information regarding anchors and destination as parameters and, upon user activation, pops up a menu with labels associated to destination anchors; each destination anchor corresponds to the URL address to be loaded in case of selection.

Pop-up Link. A temporary pop-up link can specify what text is to be presented to the user by presenting the text explicitily. In this case, the Java applet responds to the link selection by popping up a window containing the given text.

Previewing the Destination of a Link. When compared to the pop-up link, the specification for the Previewing Link needs one extra parameter: the address for the URL that is be loaded for the cases where the user chooses to complete the selection for the link.

4. Final Remarks

The applets presented in this paper provide a low-cost way for the application designer to obtain some of the features of previous hypertext systems without waiting for vendors and consortia to re-discover them. As the performance of Java implementations improves and the ability of applets to control more of the browser is provided, it is reasonable to expect that more powerful hypertext functionality will be available as add-on client functionality. The desirability of such customizibility of the hypertext system by application designers and end users was one of the key lessons of NoteCards [Halasz 88], which used Lisp instead of Java, and is still true today. Work is being carried out on the implementation of Java code supporting the linking structures specified above and examples are available in [Pimentel & Buford 96].

5. References

[Buford & Rutledge 96] Buford, J. & Rutledge, L. (1996). Third Generation Distributed Hypermedia Systems. Handbook of Multimedia Information Management.

[Carr 96] L. Carr et al. (1996) Open Information Services. Proc. Fifth Intl. WWW Conference.

[Halasz 88] Halasz, F. (1988). Seven Issues for Next Generation Hypermedia. CACM, July.

[Ladd 95] Ladd, B & Capps, M & David-Stotts, P. & Furuta, R. (1995). Multi-Head Multi-Tail, Mosaic. Proc. Fourth Intl. WWW Conf.

[Maurer 94] Maurer, H. (1994) Advancing the Ideas of the World Wide Web. Proc. Distr. Multimedia Systems and Applications.

[Meyrowitz 86] Meyrowitz, N. (1986). Intermedia: The architecture and construction of an object-oriented hypermedia system and applications framework. Proc. Conf. OOPSLA.

[Pimentel & Buford 96] Pimentel, M.G. & Buford, J. (1996). Non-Invasive Improvements to Hyperlinking with HTM (extended version). http://www.icmsc.sc.usp.br/~mgp/webnet96ext.html

[Pimentel 94] Pimentel, M.G. (1994). Evaluation of alternative operations for browsing hypertext. People and Computers IX, Cambridge.

[Shneiderman 87] Shneiderman, B. (1987). User interface design for HyperTies. Proc. Hypertext'87.

Acknowledgements

The first author acknowledges CNPq and the ProTeM-CC programme, Brazil, for the support given to the SMmD Project, and Ciro Trindade for contributions to the implementation of the applets.

WWW Technology and New Models of Learning

Mr. Jyrki Pulkkinen
Project Researcher
Faculty of Education
University of Oulu, Finland.
E-mail: jyrki.pulkkinen@oulu.fi

Mr. Esa Niemi
Project Manager
Continuing Education Centre
University of Oulu, Finland.
E-mail: esa.niemi@oyt.oulu.fi

Introduction

An important challenge facing the development of educational programs has been to find the relation between the contents and the method of study. This model of thinking has traditions which date back to the remote past. John Dewey, the American philosopher and researcher, crystallized the principle of his "problem method" in the early 20th century in the concept of "Learning by Doing". Dewey thus thought that study meant problem-solving that would be most successful through practical research activities. The same model of action is also present in the "project method" introduced by Kilpatrick. Today these principles have re-emerged as educational challenges thanks to the so-called constructive psychology, in which the basic idea is that the student must construct his knowledge and skills though his own experiences. Systemic psychology has also placed emphasis on independent research as the cornerstone of all study, but it has emphasized the social nature of learning to a larger extent than constructive psychology.

The problem in Web based courses has been the lack of functional, practical models and tools for the construction of the students' own knowledge. Just reading does not open the contents of courses, as the participants need to be allowed to do things and produce knowledge themselves. The student who is struggling alone with a bunch of WWW learning materials may well be longing for the strained atmosphere in the classrooms of his childhood. In our developmental work, an effort has been made to construct a new kind of open learning environment on Web which enables distance learning from work or home, adds the social dimension to study and helps the student to develop his thinking further.

Pedagogical principles of Open Learning Environment (OLE)

In learning theoretical or practical skills mere experience adds to the student's knowledge level but is not sufficient, however, in order to reach for higher learning processes. This would presuppose so called reflection, which can be described as a general term for those intellectual and affective actions with which an individual handles his experiences in search of a new kind of understanding and evaluation. As far as reflection is concerned, it should be remembered that although an individual controls the reflection process himself, he needs also social interaction for reflecting and comparing own ideas with other's. Individual learner can process his own experiences and link new ideas and insights into others own cognitive schemes. Teacher's role is to assist and support this process.

A reflective process cannot develop in a void. We stop to reflect because some things require reflection before action or a particular situation we are in calls for a need to reflect. Situations like this can occur in everyday life, but they can also be situations which have been created on purpose, like in our Open Learning Environment (OLE). Project learning situations arise practical or theoretical problems related to student's own learning goals, earlier experiences or related learning materials. For reflecting these problems, students can use the questions presented in the learning materials or "bookmarks" as a starting-point. Special tools for reflection and discussions are created on this purpose in our OLE. Reflective discussions are very important in order to make the students feel that they belong to the group which has come together to try and learn, to find a deeper understanding to the matter they are learning about.

Learning through own experience and reflection emphasizes the student's own activity to a great extent. The guiding principle of our OLE is that the studies proceed as a project from the student's point of view. This means that the student's first task is to get oriented towards the field of topics to be learned, and to *determine a goal* for his studies which is sensible and challenging from his own viewpoint. The goal is reached gradually through *project work* supported by different *learning materials* and joint discussions. In this practico-theoretical project work, students can put together all the knowledge that they have experienced meaningful in the learning materials and discussions.

Because the project work will be finished slowly as students proceed in their studies, it has been useful to develop so-called "portfolio study" for organization and joint evaluation of knowledge. Students have to gather any thoughts directly provoked by each section of learning materials to the relevant portfolios. On this basis it is possible to enter into discussion with the other students and with the tutor. Through this reflective and social activity on the web, the students are able to finish their project papers trying to improve their design and expertise.

Project Tools for Learning (Proto)

The backbone for the OLE and it's tools is provided by a study server (http://edtech.oulu.fi) on which information systems have been constructed to support these studies, especially project study. This information server is based on World Wide Web technology and also provides E-mail and capabilities. The learning materials and learning tools are implemented in HTML format and Allaire's Cold Fusion. This will offer quite a robust system, because all database interactions are encapsulated in a single database processing engine.

One tool for project learning consists of a remote editor for portfolios and project pages and special discussion areas for reflective discussions. The remote editor system enables the students to produce their project pages easily on the web, even without knowledge on how to use the HTML language. At the moment it will be sufficient to mark only links and images with HTML tags. In addition, students can send HTML page scripts made by another editor to our server. On the Proto pages there is also a possibility to add links to multimedia applications created by using Toolbook. To achieve system security, students have access to edit only their own pages, but they can read all the pages of other students. The tutors and administrators have access to edit all the pages. That way they can assist the students in their work, for example.

In addition to traditional information presentation, it is possible for the students to take part in discussions based on their project papers and other learning experiences. The discussion tools have been developed to support argumentation and reflective discussion. The progress of discussion is shown as a thread and students can use filters while browsing the messages. Discussion tools are also a field where the students will get assistance from the tutor, both in technical as well as in content problems.

References

Boud, D, Keogh, R and Walker, D. (ed).(1987)Reflection: turning experience into learning. London: Kogan Page.

Jonassen, D.H. (1994). Thinking Technology. Toward a Constructivist Design Model. Educational Technology April/1994, 34-37.

Järvilehto, T. (1994) Learning and the new educational technology. In Proc. Int. Workshop on CLCE, ed. by Levonen, J.J. and Tukiainen, M.T., 58-60.

Pulkkinen, J. (1996). Technology Changing the Pedagogical Culture. The Thirteenth International Conference on Technology and Education, ICTE, March, 17-20, 1996. New Orleans, USA

Pulkkinen, J. & Niemi, E. (1996). Kohti sisällön ja muodon integraatiota - koulutusteknologia 15 ov. Koulutusteknologia-lehti 1/96, 14-15.

Pulkkinen, J. (ed.). (1996).Introduction to Telematics in Teacher Training: First flexible learning material package of T3.

Sweefers, W. (1994). Multimedia Electronic Tools for Learning. Educational Technology May-June/-94, 47-52

World Band: K-12 Music Education on the Net Collaborative Learning Environments

David Reider
Scientist, Education Technologies
Bolt Beranek and Newman, Inc
BBN Corporation
70 Fawcett St.
Cambridge, MA 02138
617-873-3795
dreider@bbn.com

World Band is a project focusing on bringing new methods to K-12 music education using Internet and midi music technology. Both mediums have tremendous yet untapped pedagogical potential. Research over the past two years has resulted in new learning modes and access to previously unattainable music education concepts, especially in younger ages. The project involves member classes at geographically disparate sites who use the Internet to trade musical compositions and fragments, eventually building full works. By doing so, a new type of collaborative learning model evolves; students at very young ages with little or no previous formal musical education begin to compose and most importantly, they examine the musical process from an aural perspective rather than an intellectual or written one.

Advanced musical topics such as orchestration, form, and arrangement are studied by middle and high school students regardless of performance abilities. The work has proven that far more students are musically "aware," that is, intuitively knowledgeable, than just those that have been classified "musically literate" by traditional standards.

In a typical World Band scenario, a course that lasts a semester, exercises in the form of computer midi files are sent to students and they are guided to re-work the pieces from the compositional level. They examine elements of composition and form. Students study rhythm, bass lines, harmony, melodic shape, and timbre within a completely wholesome environment that is non-intimidating and fosters experimentation. They need only common computer editing skills and their ears. As many subject areas embrace creativity in the K-12 years (language arts, painting, drama, etc.), musical composition is often excluded because of the technical and theoretical knowledge base required by the traditional model of music education. The electronic environment permits students to hear their works as they shape them both aurally and graphically. The repertoire can be chosen by the leader (teacher), or students may work from scratch. By emphasizing composition, the student gains creative ownership and insight, and inquiry into the "how" of music follows naturally.

Students send "works-in-progress" to other connected sites, where their peers, or partners, collaborate in building the piece. Different parts of each composition are created by different composers, while relationships between students from different locations develop entirely through and within the medium of musical discourse. Musical issues, as well as ethnic, cultural, and sociological develop. This approach serves as the basis for building networked learning communities, where any subject matter may be taught collaboratively.

The presentation will discuss and demonstrate current and past projects, including a live interactive Internet concert that took place in May 1996 between students in Seoul, Korea and Stuttgart, Germany. Using a new software tool *Interplay,* developed by BBN, students demonstrated the uses, both educationally and musically, of such technology. Students performed synchronously with each other over the Internet compositions they had collaboratively composed during the previous months.

World Band projects involve middle and high schools in seven states as well as sites in Germany, Korea, Italy, and Japan. World Band has just begun a two year project with the Memphis City Schools where elementary school students from 6 schools working with high school musicians will be using new electronic music technology and the Web to study and produce a recording of blues music. World Band will enable them to share their research and production with other sites as they develop the disc.

Browsing 3D Bookmarks in BED

Serge Rezzonico (rezzoni@lig.di.epfl.ch)
Daniel Thalmann

Computer Graphics Lab, Swiss Federal Institute of Technology
LIG-DI-EPFL 1015 Lausanne, Switzerland

Abstract

In this paper BED, a prototype of a 3D bookmarks generator, is presented. This work tries to show a way of improving navigation in the Web using a cyberspatial[*] representation based on VRML. First, the general problem of creating a visual representation of abstract data is discussed and existing systems are reviewed. Then, some 3D space characteristics are explained and finally our generator is introduced and some examples shown.

Keywords: VRML, abstract data space, bookmarks, automatic generator

---

1 Introduction

Why do people get lost while moving through the World Wide Web ? Unlike a real spider web, the net millions of people are reported to be using every day for work and entertainment doesn't have a regular and beautiful structure. Like a real spider web, the net is a trap for the unprepared creatures that wander into it. Our Web doesn't have apparent limits, is not regular and is too big to be looked at all at once. One reason for this complexity is that there's no one single spider building and maintaining it. Instead, there's a myriad of single entities building their own sites with countless links to others. Naturally, when a situation is complex, humans tend to take notes, draw maps or build models. Exploring the web, one can take note of interesting places and build a list of "hot links". Why not create a 3D scene and navigate through all these links with VR techniques ? Three-dimensional navigation is considered as intuitive because of the natural human skill of moving in a 3D space. After all, that's where we live (even if we mostly move in 2D, due to gravitational constraints).

2 Creating 3D representations of data

The problem is that a 3D scene has to be generated from data which a priori has no correspondence with real physical space. Any spatial configuration that we will choose will be arbitrary. Nevertheless, this is not new. A similar problem appears when dealing with a file system and trying to give a user-friendly view of it. A huge and complex file system can sometimes be as hard to examine and search as a city's sewage system, if done in the archaic command line mode. Now that the 2D desktop metaphor is adopted by everybody, even PCs, this is an accepted fact. Some projects go one step further and develop a 3D file system representation. Because files are organized in a hierarchical way, the 3D space correlation is less arbitrary than in the general case.

Cone-Trees
One of the pioneer research groups, Xerox PARC, implemented this information visualization system where directories are situated at the apex of a cone and files and subdirectories are on the circular base of the cone. Any data element can be selected, the cone-tree then rotates bringing that element to the foreground for easier inspection. This system allows to display lots of information without becoming unusable. "Perspective walls" can also be used to show projections of the cone-tree on different planes (like cast shadows).

FSN
Developed at SGI, the "Fusion" File System Navigator [1] [Stra92] visualizes files and directories as 3D bars and platforms connected with paths. A bar's dimension and color indicates the size and age of that file.

The effort to create advanced visualization systems where the user can "read" the information faster (and possibly understand

it better) is not limited to file systems. According to [Benf95], creating a graphic representation of generic data can be done in four ways:
Data attributes can be mapped onto spatial and visual dimensions, which are the extrinsic and intrinsic dimension described by Benedikt (see section 3).
Data can be clustered with statistical methods that can show some logical similarity.
Data can be organized with a hyperstructure approach, with tree and network drawing schemes. This works well with data that are already hierarchically structured like file systems (FSN and cone-trees belong to this category).
Data can be represented with real-world metaphors, like the 2D desktop metaphor popularized by Apple with the MacOS or the recent document visualizer developed by Tenet and Webmaster in VRServer [2]: 3D corridors and doors lead to directories and files.

Q-Pit, VR-Vibe, VR-Mapper
These prototypes, developed within the **DIVE** system [Carl93], show three different approaches to abstract data visualization systems. All of them are in fact **PITS** (Populated Information Terrains), allowing to share the data with other users and be aware of their actions. Q-Pit follows the first of the four approaches just mentioned and uses an attribute mapping script to define how the intrinsic and extrinsic dimensions are used. VR-Vibe uses the statistical method and offers special "points of interest" when the user makes a query. Finally, VR-Mapper implements the third approach, generating a hyperstructure reminiscent of a cone-tree.

VizNet
Developed at the Institute of System Science in Singapore, VizNet [Fair93] is a multimedia visualization system for different types of data, from single complex objects to large collections of objects. The user can choose appropriate visualization models depending on different degrees of interest, for instance a sphere representation shows the inspected information in the center of the sphere and related information radially around it. Size, color and other visual attributes are used to encode particular proprieties of the inspected data.

Hyperbolic Visualization
This system [Munz95], a variant of the cone-tree, takes advantage of non-Euclidean hyperbolic geometry to allow the display of both large structures and detail at the same time.

3 Legibility and cyberspace

Studies in urban planning show that a city can be easier to live in when some characteristic structures are present. Kevin Lynch found more than thirty years ago that there are five important features contributing to the legibility of a city: districts, landmarks, paths, nodes and edges. The legibility of an urban environment, or generally a space, means the ease with which one can build a cognitive map of it in order to navigate through it. Researchers at the University of Nottingham used this in a prototype system called **LEADS**, which improves the legibility of abstract data spaces [Ingr95]. As the authors point out, the World Wide Web is a typical abstract data space and its legibility definitely affects the navigation. Describing the dimensionality of cyberspace, Benedikt [Bene91] introduces the concept of extrinsic and intrinsic dimensions. The first are the classic x,y,z spatial dimensions, the second are attributes like shape, size, color or spin which are logically independent of the spatial position. By carefully choosing the extrinsic attributes, the amount of information in an abstract data space can be increased without relying on text annotations or external legends, which could reduce the legibility (although legibility is related with text in the common sense).

---

4 Translating bookmarks into VRML

Netscape
One of the most popular web browsers, Netscape Navigator, automatically generates a bookmarks file. When a user exploring the net wants to take note of a particular URL[*] , the Netscape browser allows to add this address in the bookmarks list for future reference. This way, anybody can create a personal selection of the web. The bookmarks can later be selected from a pull-down menu, but this quickly becomes unpractical as the list grows. One way to solve this problem is to define submenus; even so, other information included in the bookmarks file are harder to access. A 3D representation of this file could empower the user with more information and easier navigation.

VRML
VRML (Virtual Reality Modeling Language) [3] is a standard created in 1994 for a platform-independent exchange of 3D worlds over the net. Based on Silicon Graphics' Inventor, VRML can be used to define a tree structure where every node is a geometric primitive, an attribute (like material or texture), a light source or a geometric transformation. Moreover, a geometric primitive can be associated with a URL, enabling users to make links to other 3D worlds but also anything else on the web, like HTML sites.

3D Bookmarks Generator
Our generator, BED (Bookmarks Exploring Dabbler), reads a netscape bookmarks file, parses it and generates a VRML file. A VRML browser can be used to visualize the scene and navigate through it. The choice of generating a VRML script allows

to work over the network in a platform-independent way, but also limits the kind of possible interaction to that allowed by VRML browsers (see section 7).

5 Data representation and navigation hints

Using Benedikt's terminology, the bookmarks file information is mapped onto the intrinsic and extrinsic dimensions of our cyberspace. We will refer to a bookmarks file entry as "link".

Extrinsic dimensions
A link's position in space depends on its position in the bookmarks menu structure. Groups of links are always situated inside big globes, which have a similar role as submenus or districts as explained in [Ingr95]. For example, if the bookmarks are organized in six submenus, there will be six globes and each of them will contain the corresponding links. Also, because the maximal size of a globes is fixed, a menu can be represented with more than one globe if it has many entries. All globes are situated in circular configuration, like files in a cone-tree level. At the moment, the position of a link inside a globe is not significative.

Circular globe configuration. Single links are not visible from this viewpoint.

Intrinsic dimensions
Some of the features we can use to improve legibility are: size, color, brightness, distance and shape. Others like sound and spin aren't supported in VRML version 1.0. Textures are optionally used as an esthetic enhancement, because most platforms and browsers can't assure a decent frame rate. Two intrinsic dimensions are used together to map a link's attribute. Using two intrinsic dimensions for one information (for example brightness and size for the age) further improves the legibility. The bookmarks file, other than the name and the URL of a link, stores information about its date of creation and last access. It is also very easy to get the type of link from the URL: the first part indicates if it's http, ftp or else. Here's how all this information is mapped:

Color and shape are used to represent the type of link, for instance:
blue cube: http
green sphere: ftp
yellow cone: gopher
red cylinder: telnet

Brightness and size are used to show the age of a link (last access)
small and dark: old links
big and bright: freshly accessed links.
As the background is black, big bright objects are more visible.

Some links with different size, brightness, color and shape

If textures are used, a special one marks VRML links.
Text is only used to label links and globes with their names and is pasted on both sides of a rectangular panel below each link (this way it can be read from two sides).

Navigation hints
A path connects all globes, guiding the unconstrained navigation allowed by most VRML browsers. Landmarks are present in form of stars. These stars have two roles: that of usual landmarks, marking a special point in the data space, and that of quick navigation methods, allowing the user to smoothly move to a location by selecting the star (like by clicking on it with the mouse).

The big central landmark

VRML
Some particular VRML features are used, for example levels of detail (LOD) to hide links inside a globe when seen from far outside and to draw text and textures only when close enough, to reduce rendering time. Different points of view (cameras) are used for the quick navigation mode.

From inside a globe, we see its links and also the exterior: the globes are transparent.

6 Web access

The 3D generator is accessible through the web and anybody can use it to generate a personal VRML scene. This is done with a CGI[*] script that calls the generator with the user's bookmarks URL as input. If the users configuration is correctly set, a VRML browser is automatically run with the generated 3d scene.This is the URL for the generator:

http://ligwww.epfl.ch/~rezzoni/VG/tform.html

7 Conclusion

We have shown a practical use of VRML to represent a portion of the web in 3D. This system is of course experimental and doesn't pretend to be a panacea for web navigation, in some cases using a pull-down menu is still faster. Even so, some issues of abstract data space visualization have been addressed and a few propositions that improve navigation and legibility have been made .

Limitations
With VRML version 1.0[*] only static scenes can be created. The interaction in such scenes is limited to navigation and link selection, there is no way to manipulate objects and modify the scene. Because of this, it is impossible for instance to modify the position of a bookmark interactively. Also, the generator has to be run each time a new updated version of 3D bookmarks is wanted. There's no automatic coherence synchronization between the original bookmarks file and the VRML version.

This work is part of the "Web over ATM" project (EPFL, Switzerland).

8 References

[Bene91] Benedikt, "Cyberspace, first steps", Cambridge: MIT Press, 1991
[Benf95] Benford & al., "Networked Virtual Reality and Cooperative Work" Presence (4/4), MIT Press
[Carl93] Carlsson, Hagsand, "DIVE - a platform for multi-user virtual environments", Computers and Graphics, 17(6), 1993
[Fair93] Fairchild & al., "Dynamic Fisheye information visualizations" Virtual Reality Systems, London: Academic Press, 1993
[Ingr95] Ingram and Benford, "Improving the Legibilty of Virtual Environments", Virtual Environments '95, Springer Computer Science http://www.crg.cs.nott.ac.uk/crg/Research/leads/
[Munz95] Munzner, Burchard, "Visualizing the Structure of the World Wide Web in 3D Hyperbolic Space", 1995 http://www.geom.umn.edu:80/docs/research/webviz
[Stra92] Strasnick, Tesler, "3D information landscapes", 1992
[Suth65] Sutherland, "The Ultimate Display", 1965

[1] FSN http://www-europe.sgi.com/Fun/free/cool_sw_01.html
[2] VRServer http://www.webmaster.com/
[3] VRML http://sdsc.edu/vrml/

Notes

*

The word cyberspace belongs to the group derived from cybernetics, which comes from the Greek "kybernetes", meaning pilot or governor. Cyberspace could therefore literally mean a space that can be controlled. This comes very close to Sutherland's "Ultimate Display" [1], which is a room where a computer (and thus the user) can control everything, like the existence of matter. The world cybernetics means the comparative study of automatic control systems.

*

A Uniform Resource Locator (URL) is a compact representation of the location and access method for a resource available via the Internet.

*

The Common Gateway Interface, is a standard for external programs to interface with information servers on the web.

*

Now that VRML version 2.0 is out, new features allow to create less static worlds, with more animation and interaction.

SEEWeb - An Internet Solution

Steven J. RheintgenAdvanced Software Technology Lockheed Martin Tactical Defense
SystemsUSAsteven.j.rheintgen@lmco.com

The Advanced Software Technology (AST) group of Lockheed Martin Tactical Defense Systems has developed a tailorable web based environment known as the SEEWeb. The SEEWeb provides access to an organization's Software Engineering Environment (SEE) through the World Wide Web. This paper delves further into the SEEWeb psychic by exploring the problems it solves, the technical mechanisms by which the problem is solved and how the SEEWeb is being employed to support real world projects.The problem is how to provide a widely diverse and geographically dispersed group of engineers, managers, and customers with accurate knowledge required to develop and evolve software, tailored to each user's unique perspective, and provided it in a cost effective manner. This problem is especially acute in stream-lined acquisition projects, where the Government is looking for less burdensome yet effective approaches to oversight. Traditional hard copy documentation and reporting practices are labor intensive, expensive and produce materials that are often outdated sometimes as they are even completed.Confronted with this problem, the Advanced Software Technology (AST) group developed the SEEWeb. The high-bandwidth electronic SEEWeb interface replaces the traditional, expensive, paper-based customer-contractor interface and provides both customer and contractor management with direct visibility into the state of a system or software engineering project.Conceptually, the SEEWeb is organized into three high-level areas as shown in Figure 1: Project, Product and Process. Project area contains knowledge related to project management, such as, plans and schedules. Process area shows the mechanism used to develop products. Product area contains the collection customer required project outputs. Figure 1 represents the starting point for SEEWeb navigation. The key building blocks of the SEEWeb are the ubiquitous World Wide Web and an organization's SEE. A user, with an engineering, management, or customer perspective, that has access to a browser, can dynamically obtain the latest and most accurate knowledge for a software project.The SEEWeb strives to present knowledge where existing SEEs provide information and CASE tools only provide data. For example, a requirements management tool can produce the data: 3452 requirements in the database at the moment. A SEE with integrated tools might produce the additional information that there were 3426 requirements in the database last Friday and 14 requirements have been added in the last week. SEEWebs will present knowledge:

The requirements are stabilizing. The rate at which new requirements are added has decreased over the last three weeks. The rate at which requirements are being consolidated has also decreased over the last week. Based on staff activities and historical data, we believe that the trend we're experiencing is an indicator that the requirements are well-enough understood and captured to create a baseline so that we can move on to the next phase.

This kind of knowledge would most likely be placed in the SEEWeb as an annotation by a project member. A less comprehensive version might eventually be derived from the SEE, through the use of automated mechanisms to gather metrics and analyze them in the context of historical information.

The browser provides platform independence. The knowledge being presented to the user is extracted and synthesized from the SEE in real-time. This electronic approach provides a single mechanism for intersection between these viewpoints getting total visibility into up-to-date process-, project-, and product-level knowledge and the areas of intersection between these viewpoints.

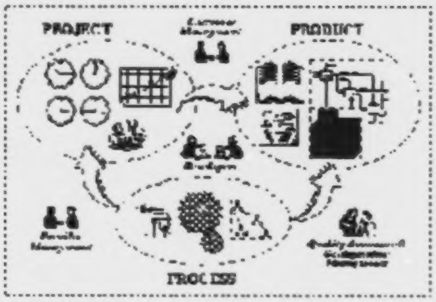


Figure 1: The SEEWeb Conceptual View

Figure 2 illustrates the flow of data produced by SEE users (managers and engineers) to a SEEWeb where it is additionally available to customers. Figure 2 shows SEE users interacting with CASE tools that store their data somewhere in the SEE. Tools may use off-the-shelf databases or proprietary persistent storage solutions. In either case, data is extracted from the databases via filters. The filters generate HTML and other data that form the SEEWeb. SEE users would continue to use the SEE to perform engineering tasks while the SEEWeb extracts and synthesizes the data from the SEE to present to the SEEWeb users.

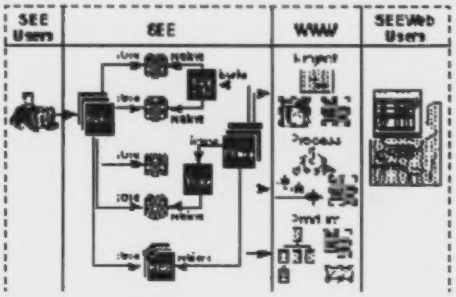


Figure 2: SEEWeb Knowledge Flow to the User

The technology base of the SEEWeb is composed of HTML and CGI bin programs. As technology evolves, the SEEWeb is continually being evaluated and upgraded with technology, such as, Java and Intelligent Agents for intelligent information integration. The SEEWeb concept is continuously refined by applying the SEEWeb to real world projects.

SEEWeb is being used to support an on-going real world project that is using a stream-lined acquisition approach to system development. The SEEWeb communicates specific critical knowledge to internal and external customers so key decisions can be made more quickly and accurate without wading through unnecessary information or raw data. The project is using the SEEWeb to convey knowledge in the form of graphics and addresses the number of software components completed, status on logged Corrective Action Reports, that is, software trouble reports. In addition, Corrective Action Reports can be submitted through the SEEWeb.

The SEEWeb provides a WWW-based interface to the information in a software engineering environment. The purpose of this interface is to provide a "window" on management and engineering activities that makes useful knowledge out of the data created by a project. The contribution of SEEWeb is the ability to present information to humans so that they can interpret the results of project activities in the light of project, product, and process knowledge.

BEST COPY AVAILABLE

Archive.edu: A Storehouse of Classroom Resources on the Web

Bernard Robin
Sara McNeil
Department of Curriculum and Instruction
University of Houston, USA
brobin@uh.edu
smcneil@uh.edu

The Need for an Educational Archive

As the World Wide Web has expanded, a number of trends have become evident. One trend is that new Web sites appear almost daily, and this ever growing list of sites vie for the attention of Web users. Although the proliferation of Web pages adds to the number of matches that Web search engines return, this volume of material makes it difficult and often frustrating for users to easily locate the information they are seeking. Additionally, we are seeing a rapid rise in the number of commercial Web sites, where copyrighted material is displayed, but users are not allowed to download and use these materials for their own purposes. Further, the ability of users to download material is becoming faster and easier as the Web evolves into a more user-friendly resource. Perhaps most significantly, the Web is now capable of distributing multimedia materials that go beyond just text and graphics. Hypermedia projects developed with software such as PowerPoint, HyperStudio, and Authorware are increasingly being disseminated through the Web.

These trends are also having a dramatic effect on the educational community as more schools and universities become connected to the Web. Many educators are wanting to, and in many cases expected to, integrate technology into their classroom instruction. The Web, with all of its potential to deliver educational materials, is often a confusing and underutilized resource for teachers and students. The current state of the Web with respect to education may be summarized as follows:

1. There is a comparatively small amount of relevant downloadable courseware, instructional graphics, video clips, and technology tutorials available on the Web for educators.
2. What materials do exist on the Web are often difficult to find, and educators frequently do not have the required navigation and search skills to access them. Perhaps more importantly, many educators simply do not have the time necessary to search for materials that can be used in their classrooms.
3. When educators do find instructionally-relevant materials, many times these materials are copyrighted. Copyright laws for the electronic dissemination of information are still evolving, and most educators are unclear about their rights and responsibilities. According to Baer [1996], even unintentional copyright infringement is still a violation and "...the same powerful tools that make your original creations possible are also unsurpassed tools for infringing copyright" [p.163].

Description of Archive.edu

To make instructional resources that educators can use in their classrooms more accessible and downloadable, we have created Archive.edu. This Web site will serve as a storehouse of instructional materials on a variety of topics that educators can easily access. But more than just another Web site, Archive.edu is a location where educators will have permission to download any of the materials. These materials can be used as they are, or customized and incorporated into other instructional projects.

The design and development of Archive.edu is based upon interviews with educators and students who have expressed an interest in locating educational materials on the Web and plan to use this resource. The prospective audience for Archive.edu includes preservice and inservice K-12 teachers and technology coordinators, university faculty and students, and others interested in educational materials.

Archive.edu is also a place where educators can submit resources that they and their students create and are willing to share with others in the educational community. We will begin this process by "seeding" the archive with a number of useful materials that have been developed by faculty and students at the University of Houston's College of Education. It is our hope that other educators from around the world agree that this method of sharing instructional materials is beneficial and can further expand the usefulness of the Web as a truly collaborative networked resource.

A variety of methods can be employed by educators to successfully develop multimedia materials that are free from copyright restrictions. One technique involves the use of digitizing student work to create graphic images. Teachers might have their students create their own drawings and paintings and then convert them to digital format. This can be accomplished using scanners, digital cameras, or video camcorders and inexpensive video digitizing cards available for a variety of personal computer platforms. In one example [O'Haver 1996], students in a fifth grade class in Maryland designed a Web page based on book reports of famous people. In addition to writing several paragraphs based on the book they read, the students drew pictures of the subjects of the reports. These graphics were then digitized and included on the Web page. This process allows educators and students to easily become multimedia developers and integrate customized images into their own hypermedia projects and Web pages.

Tutorials and instructional guides for developing technology resources will maintain a strong presence on Archive.edu. One of the consistent findings of our interviews with prospective users of the archive indicates that educators would like to have access to materials that will help them become more proficient with educational technologies [Pinell-Joffrion 1996]. Instructions for how to integrate technological resources into the classroom was also of interest to potential users.

Archive.edu has begun with the following areas or "vaults:"

- **Courseware Archive for Educators**
  For example, visitors to Archive.edu can find and download a PowerPoint presentation on Graphic Design Principles.
- **Instructional Graphics Archive for Educators**
  For example, in this area, Web browsers will find and download student-created maps, diagrams and charts as well as digital photo collections.
- **Technology Tutorial Archive for Educators**
  For example, Archive.edu browsers can find and download a presentation about how laserdiscs can be used in the classroom or step-by-step instructions on how to add sound to a hypermedia program.
- **Digitized Video Clip Archive for Educators**
  For example, users will be able to find and download short video clips of interviews with professionals who use technology in their jobs.
- **Archive of Technology-related Research**
  For example, browsers can find and download master's theses, doctoral dissertations, research papers, and works in progress related to instructional uses of new technologies.

Within each vault, materials are organized by both content area and level of intended audience. For example, the Instructional Graphics Archive is divided into individual sections for K-12 educators and higher education faculty as well as by topic. Expansion of Archive.edu is planned to include resources designed specifically for novice computer users and educators just beginning to explore ways of integrating technology into their instruction.

Educators are invited to submit their own instructional materials and that of their students to Archive.edu to share with other technology-using educators. Visitors to the Web site may find guidelines for submitting their materials at: http://www.coe.uh.edu/archive

References

[Baer 1996] Baer, M. (1996). Copyright and the visual arts. Macworld, 13 (10), 163-167.

[O'Haver 1995] O'Haver, M. (1995). Famous Americans biography book reports. [On-line]. Available: http://www.inform.umd.edu:8080/UMS+State/UMD-Projects/MCTP/Technology/School_WWW_Pages

[Pinell-Joffrion 1996] Pinell-Joffrion, C. (1996). Survey of computer-using educators. Unpublished master's thesis, University of Houston, Houston, Texas.

Description and Evaluation of the Web-Based Course, "Exploring the Internet," at the University of West Florida

Ed Rodgers, Chair
Department of Computer Science
University of West Florida
Pensacola, FL 32514
erodgers@uwf.cc.uwf.edu

Background

Since the fall semester of 1995 over 500 students have completed the Web Based course, "Exploring the Internet," at the University of West Florida. The course, taught entirely on the Internet, was developed by faculty and students from the Department of Computer Science and the Institute for Human and Machine Cognition. The University offers the three semester course for academic credit or audit credit to those who want to learn more about and effectively utilize the Internet and the World Wide Web. No prior computer coursework is required, and only a basic familiarity with a student's own system is necessary. There is no need for a student to go to the UWF campus, except to initially pickup course materials. Students access lessons and submit assignments remotely through the Internet by using software provided with the course enrollment. In order to facilitate the process, a help desk is provided to answer students' questions by telephone or email. The course has proved to be popular not only with University of West Florida students, but also with non-student members of the Pensacola, Floridda community and beyond. This paper describes the nature of the course and a critique from students who have been enrolled.

Course Description

The course introduces the student to the Internet, using the Internet itself as the delivery vehicle. The Internet is the method for submitting messages and course assignments and the source of information required to complete the assignments. Tools, such as Web browsers, mail programs, file transfer utilities, and chatting programs are presented and used. After completing the course, a student should be able to understand the nature and extent of the Internet, advantages and disadvantages of its use, and sources of Internet -based information. Through the "hands-on" approach to the course, the students who complete the course should be able to more effectively utilize the Internet and the World Wide Web.

Course Organization

The course is organized as a sequence of modules, which the students complete at their own pace. The information needed to complete each module is obtained using one or more of the available Web search engines. Assignments and tests, included in the course modules, are completed by the students and submitted to the professor through the Internet itself. Students who have difficulty may telephone or send email messages to the help desk at the University. The following modules are included in the course:

Module One

The first module provides basic information about the Internet and the means to navigate through it. Some historical perspective is provided, but the emphasis is on utilization.

Module Two

The second module covers IP addresses, URLs, and mail. It explains addressing for the Internet and the

Web and shows how it works. The module also covers how to send and receive mail using Eudora.

Module Three

The third module explores the various search engines available and provides information about how to effectively use them to conduct research on the Internet.

Module Four

The fourth module covers newsgroups and listservers. It describes how to search for newsgroups of interest, how to read and post to them, how to set up a newsreader, and how to subscribe to newsgroups or mailing lists.

Module Five

The fifth module provides information on Internet Relay Chat (IRC). It describes how to set up and use the software for talk and chat lines.

Module Six

The sixth module covers FTP. It shows how to retrieve software from the Internet.

Module Seven

The last module covers Telnet. It describes a program that allows a user to remotely login to a computer over the network.

Background and Hardware Required

One needs only to be familiar with the operating system of their computer (Windows, Macintosh, OS/2, Warp, etc.). A student should be able to install the software provided with the course and follow the instructions for accessing the courseware and submitting assignments. No knowledge of networking is required.. The specific hardware requirements vary with the computer, but it is recommended that the system have at least 4-8 Mb of RAM, 5-7 Mb of free disk space, and a 14.4-28.8 kbps modem.

Critique of the Course

"Exploring the Internet" has been a popular addition to the course offerings at the University of West Florida. A survey of students who have completed the course relvealed that most students, especially those with very little prior background, gained a great deal of knowledge and expertise and expressed a high degree of satisfaction. Many students who were already familiar with the Internet found new and exciting ways to use it and were also psositive in their comments. Only those students who were frustrated with how to use their own systems, had not adequately completed their aassignments, or were already thoroughly familiar with the material were not as satisfied.

Electronic Publishing and the Web

Henryette Roth
Journal of Molecular Modeling
Computer Chemistry Centre
University of Erlangen
Germany
roth@organik.uni-erlangen.de

Timothy Clark
Computer Chemistry Centre
Department of Organic Chemistry
University of Erlangen
Germany
clark@organik.uni-erlangen.de

The real beauty of the World Wide Web is that it is a many-to-many communication medium, where the whole world may be an extension of your mind, where some day all knowledge is accessed instantly by the electronic pulses of your brain, where the mere act of formulating a questions prompts a global search that retrieves the answer. The problem is that we loose the information about the quality of the published material, that no distinction is made between tested and untested material. [ACS 1995]

The beauty of today's molecular modeling tools is that you can play with your molecules. Simply by rotating them, zooming them in and out, looking at them from all sides, changing their shapes, it becomes easier to understand their chemistry.

However, when it comes to scientific publishing most of the beauty is lost. Colour graphics in printed journal are still very expensive. Molecules are constrained in 2D-structure diagrams. Stereo plots are poor solutions that can't help the molecules come back to life.

Electronic publishing promises the way out. The idea of presenting results in the medium in which they were created need no longer to be fantasy since the expansion of the Internet. The use of electronic media in every step of the publication process should also lead to unprecedently short publication times. The articles of a journal could be made accessible through FTP- and WWW-servers. The server-mode of publication, however, does not mean that the longevity of the published material could not be provided, this could be guaranteed by issuing CD-ROM versions. Thus the electronic journal becomes a permanent archive of science like the classical printed ones.

The nature of the computer makes innovative new techniques necessary in electronic journals in order to make this new medium truly user-friendly. Analysing molecules on your computer is great, but reading a longer text on the screen is tiring. Manipulating molecules that have been investigated by others is helpful for understanding. Changing them by accident would harm the scientific information.

Internet technology is moving fast. We can't build up today the ideal framework for electronic publishing. An electronic journal is much more a research project. We try to apply our editorial intelligence to figuring out what people really need in the online space. The goal is to develop tools that fit perfectly in today's world of scientific publishing and open doors to high technology publishing projects in the future.

The Journal of Molecular Modeling is the first electronic journal in chemistry. [Clark 1995] What started as an idea at an ACS meeting has developed into an experiment in which we are very much learning as we go along. When the Journal of Molecular Modeling was designed, we had to rely on FTP for the data transfer and Postscript as format for the papers. WWW, VRML and Java were still in their childhood and only used by some few experts. When we finally started to publish, WWW could be used to published graphical abstracts and the papers could be published in PDF (portable document format).

The development of a new journal needs the experience of a publishing house and the enthusiasm for the truly new possibilities represented by the Internet. The Springer Verlag gives us a rich source of publishing talent, while the editorial office situated at an university environment adds the creativity and enthusiasm.

The goal is to convert the Journal of Molecular Modeling, which has now in the beginning a conservative "book-like" format, into something approaching an interactive journal. Online books offer a number of unique advantages. They can accommodate a great volume of information, not only scientific papers, but also tools, 3D scenes, sample programs, videos, program inputs etc. The paper becomes the front end or interface to much larger bodies of information. The first steps have already been make. Search features make it easy to locate information not only within a single contribution, but also across the entire journal and all the related information stored behind. The incorporation of 3D-structures and videos is merely routine. The technology to incorporate further applications will be developed over the years and we will try to keep pace with its progress.

Literature References

[ACS 1995] ACS Publications (1995) Will Science Publishing Perish?.

[Clark 1995] Clark, T. (Ed.) (1995). Molecular Modeling Annual. CD-ROM and Print Archive Edition Journal of Molecular Modeling, Volume 1, 1995, Springer Verlag, Berlin Heidelberg.

Acknowledgement

The authors like to thank Peter Enders and Springer Verlag for helpful discussions and financial support.

A Web-Based Tool for Evaluating User-Computer Interaction

Jaime Sánchez and Omar Alonso

Department of Computer Science
University of Chile
Chile

Email: {jsanchez, oalonso}@dcc.uchile.cl

Introduction

Diverse studies have increasingly demanded easy-to-use and well documented tools to evaluate user interfaces [Norman & Drapper 1988; Norman 1988; Winograd 1996]. Authors agree, that whenever possible during the software design cycle, it is safe and necessary to do some sort of evaluation.

Efficient tools to register adequately data about interfaces (display, dialog boxes, controls, menus, icons) and interaction (feedback, visibility, mapping, affordances, usability) are needed in terms of having standardized measures to get a sense of how the design matches user satisfaction, needs, feelings, interests, and mapping [Norman 1988; Apple 1992].

The process of evaluation is essential to inform the designer about how well a proposed design models the users' needs in terms of their characteristics, type of activities to be done with the system, the environment of use, and the technology that supports it [Preece 1994, Barsfeld 1994, Rada 1995]. Different traditional evaluation techniques such as observation, verbal protocols, users opinions (interviews, group discussions, questionnaires), focus groups, logging actual use, and user feedback have been described in the literature in order to gather data to better fit the needs of the users' circumstances [Nielsen 1993; Schneiderman 1992].

It has been said that the process of selecting appropriate evaluation techniques involves 'picking, mixing, and adapting from the range of techniques available' [Preece 1994, Cooper 1995; Barsfeld 1993]. This study intends to introduce a web-based tool based on an interactive questionnaire to obtain users' and designers' opinion and reactions both locally and distributively. We implement a traditional technique, such as a questionnaire, as a tool that can be administered and accessed distributively, anytime, anywhere, from a Web page (see the questionnaire in http://www.dcc.uchile.cl/~oalonso/project/).

The Model

We present a model for evaluation user interaction process based on a Web application that can complement activities based on user observation. Possible users of this tool are people interested in evaluating a specific software based on a set of questions. Also, it can be useful for interface designers to experiment and collect data for a more in-depth analysis of their design. Both will be using the same software for evaluation purposes: a Web browser.

The user who wants to evaluate a particular software can interact with a Web browser as the main tool for answering questions and annotating comments. The user can check for graphical output of the data stored, select comments from other users, and visualize snapshots of other user interfaces running in other platforms.

The interface designer can check for graphical output, special comments, and other requirements. The next step after the data analysis is to rebuild or rearrange features of the user interface and place them on a Web page. This can be done in either of two ways. The designer can choose a user interface toolkit, build one, take snapshots, and place them on the page. The other approach is to build an applet and place it on the page with the advantage that the user not only can evaluate the software interface, but can also check and interact with new features while the design process is underway.

Our first testing was centered on the evaluation of Web browsers owing to the availability of different products and research prototypes, and to the fact that a Web browser can be used during the user-interface evaluation.

Implementation

The tool is a Web-based application that presents a questionnaire form to the user where s/he can choose a Web browser and then fill in the answers to all the questions. When the user submits the form, a CGI script takes all the data and stores it in a file in such a way that later can generate the appropriate graphics.

The structure of the software includes a home page with textual description of the guidelines, scores, and an example. It also presents other links to the Web sites of each browser to check for latest releases and other kind of information. Once the user chooses a Web browser to evaluate, a script generates the guidelines form on the fly. The user answers the questions and also enters some personal opinions about certain features. When the form is submitted, another script takes all the data and stores them in a logfile.

Furthermore, if a user wants to know the results, there is a link to a page with different graphics showing up-to-date data. For now, the graphics are basically histograms written in Java and the parameters for the applets are also generated on the fly [Gosling & McGilton 1995]. In a future, the idea is to have different queries to databases in order to generate customized graphics.

The construction of new features was done using both aforementioned approaches. Delphi and VisualObliq were the main tools used in the first approach and Java was used for the second approach for a simple interface.

Final Discussion

We believe that the model introduced is appropriated to perform distributed interface evaluation. This task can be performed anywhere, anytime and only requires a Web browser. From the designer point of view, the incorporation of Java applets makes the feedback process much more interactive. Users can test miniapplets in order to analyze a special feature(e.g. a dialog box). Actually, users can add applets and have more interaction with the design. Last, but not least, the pages with the guidelines, results, user feedback, and applets act as a collective memory by keeping track of the different stages in the design process.

Acknowledgments

This work was partially supported by the Chilean Science and Technology Fund (FONDECYT), Grant 1950584.

References

[Apple 1992] Apple Computer (1992). Macintosh User Interface Guidelines. New York: Addison-Wesley.

[Barsfeld 1994] Barsfeld, L.(1994). The user interface, concepts and design. New York:

Addison-Wesley.

[Cooper 1995] Cooper, A. (1995). About face. Foster City, CA: IDG Books.

[Fowler and Stanwick 1995] Fowler, S. & Stanwick, V.(1995). The GUI Style guide. Cambridge, Massachusetts: Academic Press, Inc.

[Gosling and McGilton 1995] Gosling, J. & McGilton, H.(1995). The Java Language Environment. A White Paper. Sun Microsystems, Inc. Cupertino, CA

[Nielsen 1993] Nielsen, J.(1993). Usability engineering. Cambridge, Massachusetts: Academic Press, Inc.

[Norman and Drapper 1988] Norman, D. & Drapper, S.(1988). User-centered system design. Hillsdale, NJ: Lawrence Erlbaum.

[Norman 1988] Norman, D. (1988). The psychology of everyday things. New York: Basic Books.

[Preece 1994] Preece, J.(1994). Human-computer interaction. New York: Addison-Wesley.

[Rada 1995] Rada, R.(1995) Interactive media. New York: Springer-Verlag.

[Schneiderman 1995] Schneiderman, B.(1992). Designing the User Interface. New York: Addison-Wesley.

[Winograd 1996] Winograd, T.(1996). Bringing design to software. New York: ACM Press.

Providing Real Time Instruction on the WWW: Lessons Taught and Learned

Robert Sapp
John Hopkins Univ. of Medicine, USA

Introduction

From June 24 to June 27, 1996, the William H. Welch Medical Library of the Johns Hopkins University School of Medicine implemented the course, "Current Topics in Health Sciences Librarianship." Supported by grant number 1 T15 LM07094-01 from the National Library of Medicine, the project introduced important new methods for web-based distance learning. This paper will examine the WWW functions used to support and deliver the real time instructional events used in this program. Discussion will cover the technologies used to provide the synchronous delivery of audio and HTML supporting content and the web- based applications used for instructional feedback and evaluation. Specific applications include Progressive Networks' RealAudio Encoder, Server, and Player; custom written scripts which synchronize and push HTML content; web-based forms for feedback and session evaluation; Hypernews; and additional web-based technology.

The Grant

Health sciences librarians (HSL) improve health by enhancing the availability and use of biomedical information. The long-term objective of this project is to make HSL more effective in their role by using emerging technologies to create a lifelong learning program. This program will enable HSL to update and extend their professional education and training regardless of their physical location.

The specific aims of this project are to plan, implement, and evaluate a one-year pilot continuing education (CE) program for HSL, "Current Topics in Health Sciences Librarianship." This program will offer instruction in a general-interest track, instruction in a specialized track, and a conference on related topics in parallel with these two tracks. The program will consist of four complementary and interconnected initiatives: (1) An On-site Course, (2) A Telecourse, (3) An Online Course, and (4) Electronic Proceedings. This session examines the web-based technology used to implement the online portion of the course.

Elements of Instruction

Online Course Home Page

The Course Home Page provides general information about the program as well as links to program (list and tabular format), information for participants, and the project team. Links to additional information including participants, posters, faculty, student lounge, course library, and help are also available.

Presentations

The majority of the primary instruction is provided through presentations. During the live delivery of the course, students connected to the presentations through a passworded link. Once connected, they received the synchronized audio and HTML slides in real time. At the end of each presentation, each student had an opportunity to enter questions for the presenters on a web-based form. Finally, a session evaluation form using radio buttons and textual fields was provided.

Student Lounge

The student lounge used HyperNews to facilitate a series of informal discussions based on the course content. Students were encouraged to interact with faculty and other students.

Posters

Students were required to submit electronic posters for inclusion in the program based on the five major areas of content: (1) Education, (2) Information Technology, (3) Market Forces, (4) Publishing and Copyright, and (5) Special Topic: The Informatics of the Human Genome Project. Each poster varies in format and delivery, but includes links to a HyperNews thread for discussion and to the author's email account for one-on-one interaction.

Course Library

The Course Library consists of a series of links to relevant WWW resources. Divided by major content area and session, the course library includes areas for faculty and students to suggest important links.

The Technology

Bell Atlantic Distance Learning Facility

This facility provides a distance learning environment through the use of interactive audio and video between a home site and one or more remote sites. The discourse is maintained by a bi-directional signal shared by similarly configured distance learning facilities. Instructors and students in participating classrooms are able to interact using a system of fixed and moving cameras, video monitors, omni-directional microphones, and high-fidelity speakers. This classroom acted as the initiating point for the instruction.

RealAudio

The composite audio received from the Distance Learning Facility's Audio Patch Panel was fed directly into the line-level in of the SoundBlaster card installed in a Pentium 120 running Windows 95. Here, the audio was digitized and encoded using the RealAudio Encoder Version 2.0. The encoded audio was then transported to the RealAudio Live Transfer Agent, which prepared the audio for distribution in real time, and the RealAudio Server, which provided the audio stream to students participating in the course online. Both the Live Transfer Agent and the Server were mounted on a Sparc 20.

Slides

RealAudio provides the ability to synchronize audio with HTML documents using a "Synchronized Multimedia" technology. However, since the URL's must be timed and their duration predetermined, this method is not sufficient for any real time dissemination. As noted by Progressive Networks, it is not possible to use "Synchronized Multimedia" with a live feed. Since the specification for primary instructional delivery included html accompanying the audio, we needed to pursue alternative methods for delivering the slides to students. In the end, we decided to send the URLs to the participants by using server push technology.

The 18 presenters over the course of the four day session represented all levels of technical expertise. Early on, we determined the need for an intuitive interface that provided both linear and hyper navigation tools for the slides. The presenter's interface was a Netscape 2.0 browser with two frames. The top frame, or navigation frame, was a simple interface allow the presenter to move forward, backward, reload, and jump to a specific slide number using labeled buttons. Written in javascript, these simple tools provided even the most novice presenter with a comfortable interface.

As the presenter moved through the slide show, the program captured, copied, and parsed the URLs, providing one copy appropriate for the presenter and another appropriate for the participants. The lower frame of the presenters interface displayed this parsed slide. This interface provided two additional boons: (1) all links displayed in the lower frame remained active, and (2) if the presenter choose to use these links to connect to external sites, those URLs would also be pushed to online students. The reload button in the navigation frame returned the presenter and students to the slide from which they departed. Students attending the online course saw the HTML content presented in a single frame synchronized with the RealAudio feed.

WWW Forms

Since the audio and slides were single direction, we needed to develop and implement additional tools to provide mechanisms for interaction. The final two slides from each session included two WWW forms. The first was a single free text field which sent questions and feedback to the presenter. These comments were made available on the presenters client, allow each instructor incorporate them into the interaction by reading and responding to them. The final slide was an evaluation form. Using a series of radio buttons and text fields, we built an instrument to measure students' responses to the instruction and mastery of learning objectives.

HyperNews

HyperNews combines the functionality of the web's hypertext and the communal interaction of UseNet. Using a series of topic-based hierarchies, it allows users to comment on existing "threads" and to begin new ones. We used HyperNews to create in informal venue for instructors and students to interact. More information about hyper news is available here.

Conclusion

In the 14 months since we began planning the "Current Topics in Health Sciences Librarianship" course, we've met both obstacles and possibilities in delivering live instruction on the WWW. We know now that low bandwidth, poor connectivity, non- standard software, and quirky helper applications provide real and substantive obstructions in delivering real time instruction on WWW. We know also that patient students, modest interfaces, and creative developers can solve, avoid, and account for these problems, often as they occur. In scale, the delivery of real time instructional on the WWW now possible. Let the walls fall where they may.

Professional Development On The World Wide Web: Training Secondary Mathematics Teachers

Paul G. Shotsberger

Christopher G. Spell

Mathematical Sciences Department
University of North Carolina at Wilmington
601 S. College Road, Wilmington, NC USA 28403-3297
e-mail: shots@cms.uncwil.edu, spell@cms.uncwil.edu

Background

Teachers in North Carolina are required to participate in inservice staff development as a part of a licensure renewal cycle which repeats every five years. This training normally takes the form of workshops conducted at a single site over the course of several days or even several weeks, typically during the summer months. Drawbacks to this type of staff development include the time and money required for travel to the training site; the artificiality of training done in isolation of the classroom, wherein teachers must wait until the school year begins again to apply their learning; the logistical problems of attempting to support teacher change once teachers have returned to their schools; and the difficulty inherent in trying to keep teacher training current in a discipline such as mathematics education where knowledge and practice are rapidly evolving.

We are developing a Web site for teacher training to address these shortcomings. Using this technology, professional development and support would be available to teachers at their schools during the school year, and the on-line training and resource materials could be updated on a continual basis in order to maintain teacher awareness of current trends. Further, we believe that sufficient interactivity can be built into the system so that teachers will feel that they are a part of a community of educators despite potentially being the only workshop participant at their school.

Design

The distance teacher training system we are developing is a World Wide Web site called INSTRUCT, which stands for *I*mplementing the *N*CTM *S*chool *T*eaching *R*ecommendations *U*sing *C*ollaborative *T*elecommunications (http://instruct.cms.uncwil.edu). INSTRUCT is designed to introduce secondary mathematics instructors to the National Council of Teachers of Mathematics *Professional standards for teaching mathematics* (NCTM, 1991). INSTRUCT's design integrates aspects of groupware, or software intended to support group interaction, to expand its use beyond being simply a storehouse of instructional material. INSTRUCT provides the capability for users to meet synchronously or asynchronously, as appropriate. Following is a listing of the options available through INSTRUCT with a description of their function.

1. A hypermedia version of the NCTM Standards for Teaching Mathematics found in the *Professional standards for teaching mathematics* [NCTM 1991] -- This choice links users to a Netscape page with the following menu items: Worthwhile mathematical tasks, Teacherís role in discourse, Studentís role in discourse, Tools for enhancing discourse, Learning environments, and Analysis of teaching and learning. Each of these sub-menu items link to other Web pages which employ text, images, audio and video to provide the user with a multimedia introduction to the NCTM Standards for Teaching Mathematics.

The *Professional Standards for teaching mathematics* includes vignettes intended to act as exemplars of the standards in practice. Similarly, INSTRUCT provides the user with multimedia vignettes to enhance and clarify presentation of the training material. Teachers have already been given the opportunity for involvement in the project as contributors of student work, lesson plans, pictures, and audiotape and videotape of classroom activities for building multimedia vignettes. The use of authentic classroom materials, such as written records of student problem solutions or checklists of observed problem-solving behaviors, is intended to make INSTRUCT both practical and useful for the continuing education and support of mathematics teaching.

Each sub-menu page will contain its own "Check for Understanding" form for the user to fill out and submit to the training coordinators for assessment of the userís mastery of the standard. Assignment of licensure renewal credit will be based on successful responses to questions about the standards and on teacher reports about their implementation of classroom activities intended to reflect the standards.

2. On-line Educational Resources -- Choosing this menu item links the user to a Web page of educational resources currently available on the World Wide Web, each of which in turn links users to the desired Web site. Sites are grouped by category, such as Geometry and Chaos, History of Mathematics, Lesson Planning, National Agencies and Information Sources, Statistical Data Sources, and Technology Resources. Additionally, this page provides direct links to the North Carolina Department of Public Instruction and K-12 Schools On-line.

3. Attend a meeting -- Choosing this menu item launches Netscape Chat©, connecting users to the Chat server located at our university. Chat's What You See Is What I See (WYSIWIS) characteristic allows meeting participants to synchronously communicate with each other while viewing the same Web site. The WYSIWIS feature is especially important because it allows training coordinators to conduct on-line meetings using Web pages from INSTRUCT to facilitate the discussion.

As indicated above, trainees will be given assignments to carry out in their own math classes in order to promote teacher active participation in and application of INSTRUCT training. An intentional by-product of these assignments will be to encourage the need for sustained interaction and collaboration among participants and coordinators. The meeting option would be particularly useful for INSTRUCT training coordinators to provide additional training materials and commentary to users, for trainees who wish to meet with an INSTRUCT training coordinator regarding their progress, or for users who wish to get together on-line to brainstorm about classroom ideas and plans.

4. Join in a discussion -- This option connects users to a threaded discussion area where they can asynchronously communicate with other INSTRUCT users at their own convenience. The format being described here benefits teachers by allowing them the opportunity to be involved in more long-term discussions about issues raised in the NCTM Standards, by facilitating the sharing of news and other items of interest between colleagues, and by affording users continual access to previous communications via discussion histories.

5. Send a message -- This choice provides Internet mail access for making a meeting appointment or for asking a question of a training coordinator.

Support

A long-term goal for INSTRUCT is that teachers would continue to use the program after completion of the Standards training. The menu options for on-line resources, meetings and discussions provide users with resources which can enrich classroom teaching on an ongoing basis while providing support for teacher change. Further, an essential feature of Web pages is that they are extensible and easily modified. The developers envision other training modules being added to INSTRUCT, so that it can grow and change to meet evolving teacher licensure/training needs.

Reference

[NCTM 1991] National Council of Teachers of Mathematics (1991). *Professional standards for teaching mathematics*. Reston, VA: NCTM.

Discursive Networking: Context Generation in a Networked Environment

Julean A. Simon
Simmelstr. 33, 13409 Berlin, Germany
(0049) 030 4919760
simon@berlin.snafu.de

David C. Wohlhart
Pädagogisches Institut, Graz, Austria
(0043) 0316 574541
dwohlhart@borg-6.borg-graz.ac.at

Abstract: DisNet is an environment for structured communication and collaborative working, learning and playing. It conforms with the basic idea of hyper-systems in that it allows for representing context as a linked, nonlinear knowledge structure integrating multiple data types. Additionally DisNet integrates the common capabilities of hyper-systems with features supporting the generation and dynamic differentiation of contextual relations.

"Discursive Networking" attempts to organize collaborative learning and working as an open discourse and to represent the development-process of a context as a networking of semantical units on the basis of their relations. The context which is continuously differentiated by collaborators in a group-project is accessible by them on different levels: these are represented as graphical domains of interaction, which are coupled structurally.

The pedagogical goals aim at improving argumentative preciseness, acquisition of analytical and integrative competencies and developing awareness of knowledge-relations.

Project Background

The concept of DisNet has been developed and specified by Simon & Wohlhart since 1994 as part of "Comenius", an educational, broadband-network research-project, commissioned by the German Telekom. The idea of this ongoing pilot-project is to interconnect five Berlin schools by an ATM-network and establish a suitable network-environment. Beyond this practical side the project required to develop a model of collaborative networking appropriate to pedagogical demands and to study the effects of the environment on collaborative learning. At the time of writing the network has been set up, a first version of the software has been installed and the students are about to start the project-work. This paper refers to a more recent specification of DisNet, which Simon & Wohlhart are in the process of developing. It describes a highly user-definable, multi-purpose environment for collaborative work and can be used to implement a broad scale of use cases.

Pedagogical Requirements

While bi-directionality and n2n-connectedness is supported on a low level of network-systems, many network-applications implement one-way-concepts and assign users the role of consumers. This type of personified client/server-relation is also common in the domain of educational systems and resembles the classical pedagogical situation of a teacher presenting the syllabus in front of a class. We find this provider/consumer-concept applied also in the majority of standalone learning software: albeit the "drill and practice"-approach is about to be replaced by a "guided exploration"- paradigm, the main interactional category of hyper-systems is still choice of preproduced contents. This trend is very likely to stabilize in view of the discovery of the education market by the software-industry and content-providers as well as through a redefinition of education as a chargeable service.

In view of the implications of the information-society school can be described as a place of convergence, condensation and structuring, but also of the generation and differentiation of knowledge. These processes of knowledge-acquisition have to be actively induced by a learning-community and will be vivid even more if real and dynamic connections to the cultural, economic, political and private environment are maintained. From this perspective common notions as of the transferability of knowledge, of learning as a simulation of situations, and the traditional roles of the student as a receiver, and the teacher as a mediator of established knowledge tend to become obsolete. Rather, knowledge is acquired in the communication and interplay of self-responsible actors.

The communicative structure supported by an educational network-environment has to be open to promote types of interaction that may generate the necessary collaborative dynamics; the underlying system has to establish this openness and provide transparency, configurability on several levels and adaptability to different tasks. Several dimensions of openness can be indicated to determine the structure of the domain of interaction: communicative openness implies the maxim, that all participants be equally and fully embedded in the information-flow (in practice restricted by privacy protection, or the economy of distribution); operative openness allows the participants to configure the interfaces and devise the structure of communication and cooperation in a project; strategical openness is secured in that project-oriented methods and strategies may be discussed, dynamically established, or modified in a meta-discourse.

This form of collaborative learning can be typified as an open discourse. In this context a discourse is thought of as an iterative and (self)referential structure of interactions in a problem-field: it is an iterative process where new thoughts build upon previous ones, whereby a reference structure develops; and a discourse is self-referential in that the methods and strategies to be applied, its organization and scope are also made its subject (meta-discourse). A discourse is open in that its organization is promoted by individual, as well as group-specific approaches. It is open as it may be placed into an authentic problem-domain and also in that it is interdisciplinary and couples different types of users.

Conceptual Model

The logistics of a GUI follows and mirrors the system-designers' conceptual model [Gorny, 95] of a given task and it significantly models the way the user conceives and manages this task. A discourse, or more generally a collaboration in a network-environment can be seen as the structured interaction of participants and as the process of establishing and elaborating a context. This concept deals with interacting users and messages sent between them and leads towards a model of collaboration on the basis of the physical connectivity. In this super-observer-view (classical information theory) a message sent from userA (source) can be observed to be received by userB (sink).

Seen from the participant-perspective a userA has to receive a reply from userB to "observe" (infer) userB receiving the message. Controlling the relation between some action and the feedback userA derives the concepts of and constructs what s/he interacts with: in this model a participant interacts with the context in first place.

Context emerges as the inter-text of relations of singular operations [Rezabek, 93] and can be described as a network of relations of semantically charged components and of processes of production (i.e. generation, transformation and destruction) of these. Interactions are compensatory changes in that they are bound to a previous state and balance the state changes. Through their interactions and transformations these contextual components continuously regenerate and realize the network of processes and thereby constitute the basis for the self-organization of the system.

Interaction Model

DisNet is an environment for structured communication and collaborative working, learning and playing. It conforms with the basic idea of hyper-systems (e.g. [Hyper-G], [Hammwöhner, 93], [Keep, 93] , [Kuhlen,91], [Willenborg, 94]) in that it allows for representing context as a linked, nonlinear knowledge structure integrating multiple data types. Additionally DisNet integrates the common capabilities of hyper-systems (i.e. presentation, browsing, retrieval) with features supporting the generation and dynamic differentiation of contextual relations.

DisNet stands for "Discursive Networking" and is based on the idea of representing the process of a collaboration as the development of a network of objects embodying discursive interactions: interactions accomplish differentiations of the shared contextual network and are realized as linked objects. When a participant connects a new object to the contextual network it does not only differentiate the local relation at the connection site but calls into activity all components that can be activated by it. The activity of a component is the charge it receives through its relations to the contextual ambiance and that which it discharges into the context. Participants in a networked collaboration determine the relation between components as well as the structure of their relations and thereby establish and differentiate the context and its constituents in a collaboration. No clear distinctions can be made between interactions concerning the various levels of a collaboration. This suggests to utilize a common paradigm to represent interactions of different types.

The DisNet Model

Interactions are represented as objects which have several properties, or "slots". The slots of an object may contain as their entry the object's title, some content (e.g. a statement, a picture,...), its purpose and relation to the local and global context, sources, keywords, some behavior of this object, etc. The slot-list of an object is fully user-definable. The slot of an object may itself be an object consisting of slots.

An object is always connected to at least one other (reference-) object by means of a typed link [Conklin, 88], that names the relation between them. While objects are implicitly linked to other objects by the relation of entries in corresponding slots, typed links explicitly specify the kinds of entry/entry- and entry/context-relations (Methods like search, filtering, retrieval, redundancy prevention, complexity reduction, etc. rely on the more or less explicit representation of the relations on which they should be applied). The connectedness of an object is one of its properties; in effect every object stores part of the contextual network.

The "configuration-object" which is part of a project's home-object has the same datastructure; it determines the repertory of object-properties (slot-list), the appearance and behavior of objects in the domains of interaction (scripting). An object connected to a reference object inherits its configuration; the configuration can be modified by overriding the settings.

"Container"-objects are used to aggregate components of a network and thereby structure the net hierarchically and aid in complexity-reduction. Different types of containers (clusters, paths and collections) define the relation of the contained objects. A container allows to expand and to collapse its content and substitute it by a placeholder. A special type of container, the "local context", specifies all objects of immediate contextual relevancy for a given object.

The graphical user interface supplies two main (and several auxiliary) domains of interaction: a globalDomain, that can be navigated and allows to access the contextual network on its structural level; and a localDomain where the context is accessed on the object-level by means of multi-field slot-editor, called "Jector". The localDomain is a plane in front of the globalDomain. Another domain allows to arrange field-entries on a two-dimensional working-pane and to finalize documents and printouts. The domains are coupled in that interactions in one domain yield corresponding differentiation in the other domains.

The globalDomain displays a context as a two- or three-dimensional, semantically linked graph of iconized objects and allows for browsing and modifying its structure via user-centered navigation; that is, a participant navigates in the context-space and is provided a range of possibilities to arrange and design its appearance. In the localDomain an object is represented by the Jector, an obJECT-browser and -editOR, consisting of context-sensitive popup-menus to select slots and of fields to display the corresponding entries. Fields may display heterogeneous multimedia data including text, images, audio/video, etc.(media-integration). The Jector can be collapsed to an iconized label residing on the localDomain workbench, or expanded to show several slot-entries at a time and compare them with slot-entries of other objects.

The Jector not only enables the user to view and edit object-slots and to create new and browse existing objects, but also to modify the object-configuration, and the appearance and functionality of the Jector itself.

Interaction-based Networking

The approach suspends the participant (in a mailinglist, owner of a web-page) as the outstanding point of reference and makes the confined contextual interaction the constitutive unit of a discourse, or collaboration: instead of contributing a lengthy reply/page covering his or her view on several points, the participant interacts with the contextual structure locally. In order to adapt and fine-tune an interaction specifically to the local binding site several levels of context may be taken into consideration: the definition of a project's subject and intention, the site's location in the overall context and the local context established by neighbored objects. A contextual interaction is determined also by discursive qualities such as the depth of detail, the type of language, the level of discourse, the methods and strategies defined in the project's working-conventions, etc.

Arguing in a discourse is different from arguing in a linear text and generates a different contextual structure than partitioning a linear text and arranging it in linked nodes, as done in author- and hyperware. Discursive networking is an inherently parallel and concurrent process; links may be established at any time to any object, without the need to follow linear threads. By observing the on-going modification of the network the participant can immediately spot the centers of activity. The structure of the contextual network can be differentiated by adding new objects and connections, by segmentation (forming subnets), stratification (meta-discursive layering) and second order differentiation: operative (configuring the interface and domains of interaction), or organizational differentiation (establishing working conventions and rules of collaboration).

Although this description focuses on textual contexts it should be mentioned, that the environment supports multimedia data and that the "contextual network" can also be used to represent the interaction of objects assigned particular behavior.

Networking Contextual Relations (Pedagogical Aspects)

The iterative differentiation and reinterpretation of the contextual network and the circularity of the feedback-loop of interaction is the basis for a selfregulation of the system. The basis for context generation has to be established by the "discourse culture" in a group, which can not really be learned by distinct methods: a discourse is unique (usually irreproducible), non-sequential (participants work partially in parallel and are synchronized for communication) and non-procedural (it is impossible to describe it in advance); interactions are state-dependent, based on common languages, media, procedures, and often the roles of participants are asymmetric. The DisNet-concept is based on the (constructivist) idea of seeing communication as a structural coupling in a consensual domain, which generates a synchronization of behavior through mutual adapting and orienting interactions [Maturana, 80].

The approach is in accordance with the principle of the so-called "Projekt Methode" in that the learning process is in tendence not predetermined by a final goal and a sequence of intermediary steps, but is driven by the dynamics of local interactions. It is not primarily based on the transfer of "objective knowledge", but on the local adaption and application of a participant's personal knowledge and its interference with that of co-workers. Applied knowledge does not claim general validity, but is confined to the context in question; its validity is established in the interaction of arguments. Thereby new knowledge is generated through the synthesis of local operations, the interplay of ideas and in the observation and reflection of the history of a discourse within a proprietary contextual domain. The pedagogical goals aim at improving argumentative preciseness, acquisition of analytical and integrative competencies and developing awareness of knowledge-relations. We understand that a main feature catalyzing learning is the feedback between mutually stimulating interactions, such as orienting, searching, interpreting, altering, extending, etc. The interface has been designed to motivate switching between different types of interaction and between the domains of interaction.

Besides the authority over the content, the student is given also the responsibility for the organization and maintenance of the environment, which intends to foster key-qualifications (self-responsibility, capability to work in teams, democratic and social engagement).

Summary

DisNet is an environment that supports a discourse-driven and -driving, constructive process of networked interaction; the approach attempts to extend the notion of networking as "collecting and redistributing information" towards the "collaborative generation of context". The paper reflects considerations concerning the creation of context and its continuous differentiation by the interaction of participants in a discourse or group-project. The contextualization process is fostered less by document-oriented and author-centered contributions than by the participant's local (distributed) interaction with a discussion's, or a project's context. In the environment an interaction is represented as an object which has assigned a set of properties by which it is related to other objects. A context is established by the contents of objects, but also and more than in conventional hyper-systems by implicit and explicit relations between these contextual objects. The environment provides several domains of interaction and supports different types of interaction, including the reorganization of the working-environment and thereby implements an important aspect of a discourse: to integrate in a context both, the treatment of a subject-matter as well as the continuous reflection and redefinition of the methods and strategies and the course and the level of a collaboration.

References

[Conklin, 88] Conklin, Begeman; gIBIS, a Hypertext Tool for Exploratory Policy Discussion, ACM, ToOIS, Vol6, Nr 4, Oct. 88, pp 303-331

[Gorny, 95] Gorny, Peter: EXPOSE, HCI-Counseling for user interface design, in: K. Nordby, P.H. Helmersen, D.J. Gilmore, S.A. Arnesen: Human-Computer Interaction - Interact '95, Chapman & Hall, 1995.

http://www-cg-hci.informatik.uni-oldenburg.de/EXPOSE/EXPOSE.html

[Hyper-G] Hyper-G Bibliography: http://hyperg.iicm.tu-graz.ac.at/

[Hammwöhner, 93] Hammwöhner, Rittberger, "KHS - ein offenes Hypertextsystem", in "Information Retrieval 93", UVK

[Keep, 93] Keep, McLaughlin, Robin: The Electronic Labyrinth, 1993, QSILVER.QUEENSU.CA/pub/english/hypertext

[Kuhlen,91] Kuhlen, R., "Hypertext ", Springer 1991

[Maturana, 80] Maturana H., Varela F., 1980, Autopoiesis and Cognition, Boston Studies of Science, v 42

[Rezabek, 93] Rezabek, Heath: Autologue, Interdisciplinary Dialogue with the Global Network Environment, 1993; http://www.uio.no/~mwatz/futurec/rez/autologue.dir/atologue.html

[Willenborg, 94] Willenborg,J. "Terminologiebasiertes Hypertextretrieval", in "Texte verstehen", UVK14, Konstanz, 1994

WWW3D: A 3D Multi-User Web Browser

Dave Snowdon
Dept. of Computer Science,
The Unversity of Nottingham,
UK
d.snowdon@cs.nott.ac.uk

Lennart Fahlén and Mårten Stenius
Swedish Institute of Computer Science,
SWEDEN
{lef, mst}@sics.se

1. Introduction

Most current web browsers display only the representation of the web document(s) that the user is currently examining with no indication of the structure of the document other than by highlighting hypertext links. Those browsers that do indicate the structure of web documents, such as the Harmony browser for Hyper-G [Hyper-G], use a different window from that of the document to present this information to the user. In this paper we will describe WWW3D a novel browser that provides a single 3D display which integrates the display of the web documents themselves, the structure of the part of the web that the user has browsed and history information showing the links the user has followed in the recent past. WWW3D is implemented in DIVE [Hagsand 96] and supports multiple concurrent users who are visible to each other and who may either be browsing the the same or different sets of web documents.

3. Producing 3D Representation of a Web Page

WWW3D uses the information contained in HTML tags to produce a representation of the document in 3D space. A web document is represented as a sphere which is labelled with the document's title. The contents of the document is placed around the inside surface of the sphere. Displaying large amounts of text in a satisfactory way is difficult in current VR systems so textual information is currently represented by icons that can be unfolded to reveal the entire text. The first few words of the piece of text are displayed under the icon to give some indication of the contents. Images are displayed by texture mapping them onto polygons on the inside surface of the sphere. Finally, links to other documents are represented as icons labelled with their destination. To reduce the visual complexity of the virtual environment WWW3D makes extensive use of Level of Detail (LOD) operations. When viewed from outside a document is represented as an opaque sphere and the actual document contents is not displayed. When a user enter's a document to view it the sphere is drawn in wireframe so that the rest of the virtual environment is still visible. Figure 1 shows the contents of a web document as displayed by WWW3D.

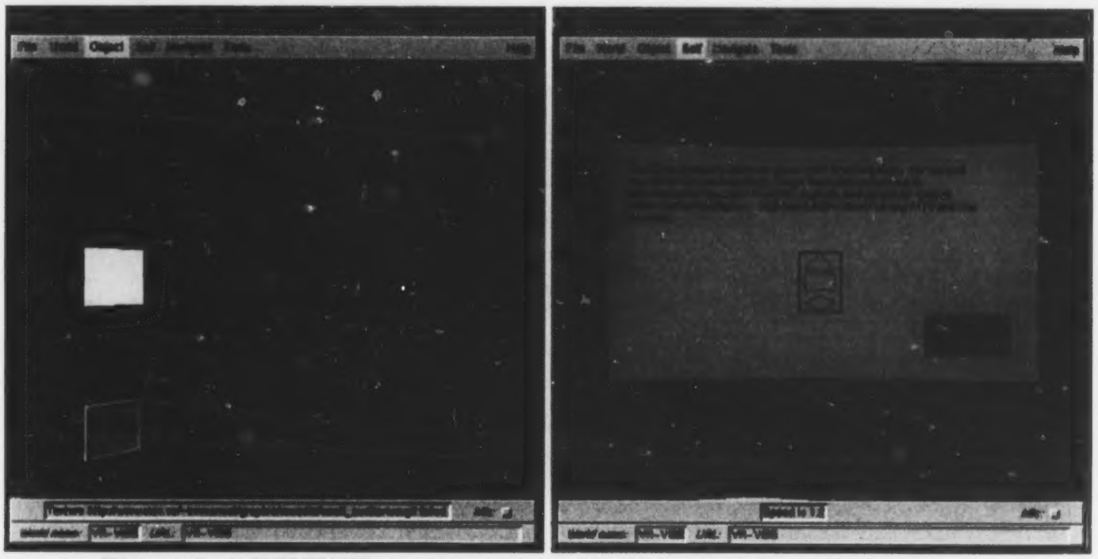


Figure 1: A WWW3D representation of the contents of a web document.

4. Browsing the Web Using WWW3D

When the user selects a link icon, WWW3D creates a new sphere representing the target document and places it near the document the from which the user selected the link. An incremental version of the Force Directed Placement (FDP) [Fruchterman 91] algorithm is used to arrange the document sphere in 3D space. In order to indicate the structure of the portion of the Web that the user has explored WWW3D draws arrows between the spheres representing linked documents. In addition to this the brightness of the arrow is dependant on the time since the user last followed that link therbey providing the user with a visual representation of their browsing history. As WWW3D parses a newly retrieved document is checks for links to documents that the user has already explored and draws arrows to represent them. This means that at any given moment the complete set of links between document is displayed without the user having to follow every link. This is intended to aid the user by indicating links between documents that the user might have been unaware of. Figure 2 shows the display produced by WWW3D after the user has browsed a number of documents. The FDP algorithm has resulted in the formation of clusters of closely linked documents. The colours of the documents provide some indication of how long ago the user last visited them.

BEST COPY AVAILABLE

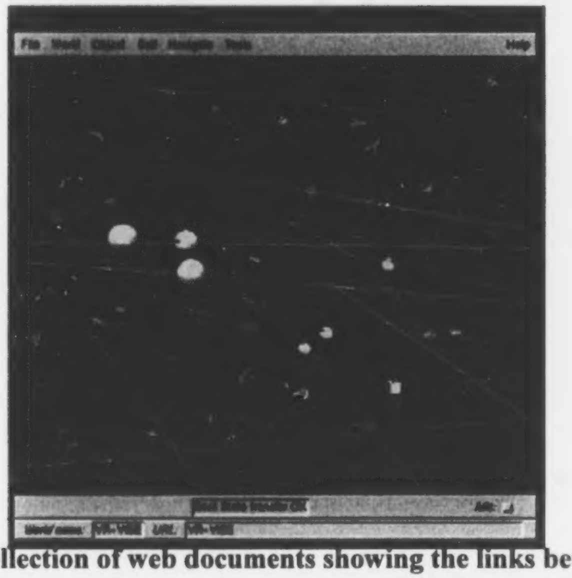


Figure 2: A collection of web documents showing the links between them.

6. Summary

In this paper we have described WWW3D, a novel Web browser that integrates information about the contents of web documents, the structure of the portion of the Web explored by the user and history information. WWW3D is implemented using the SICS DIVE VR system and can support multiple concurrent users. Users may communicate with each other using standard facilities provided by DIVE.

7. References

[Fruchterman 91] Fruchterman, T.M.J., and Reingold, E.M., (1991). Graph Drawing by Force-directed Placement. Software Practice and Experience, 21(11), 1129-1164.

[Hagsand 96] Hagsand, O. (1996). Interactive Multiuser VEs in the DIVE System. IEEE Multimedia, 3(1), 30-39.

[Hyper-G] http://www.tu-graz.ac.at/about

Acknowledgements

Many thanks to the Swedish Institute of Computer Science (SICS) for funding Dave Snowdon as a guest researcher at SICS for the period of this work. We also thank Emmanuel Frécon for help with DIVE.

BEST COPY AVAILABLE

Convergent Computer Research at the Undergraduate Level: The World Wide Web as Interdisciplinary Teaching and Learning Tool

Elizabeth Sommers, Ph.D.
Associate Professor, Department of English
San Francisco State University, USA
esommers@sfsu.edu

In this article I discuss a course I have developed at San Francisco State University (SFSU), a seminar for undergraduates entitled "Convergent Computer Research." The philosophical underpinnings of "Convergent Computer Research" are rooted in NEXA, a convergent program at SFSU that transcends the traditional boundaries between and among disciplines and scholars in the sciences and humanities. The NEXA Program reaches out to the sciences as the carriers of a mutually shared culture--what Alfred North Whitehead has termed the general form of the forms of thought--that underlies and informs the cognitive activities of apparently diverse disciplines within a common historical tradition. Many students and faculty members want to move beyond the conventional definitions and limitations of their disciplines to explore the nature of scientific innovations, ethical issues and emergent fields. The result is that NEXA continues to thrive at SFSU.

NEXA courses are deliberately convergent rather than interdisciplinary in the usual sense. That is, while some courses occasionally reveal remarkable similarities in ways of thought in the disciplines, just as often they reveal contrasts, dissensus, radically different ways of knowing, thinking, and communicating. Simply put, in these courses we attempt to stop fragmenting our ways of knowing, to melt down our classification systems. Students and faculty often remark on the satisfaction they feel having professors who represent both the sciences and the humanities, who may or may not agree, whose fields at least on first and second glance perhaps seem very different--but whose ways of thinking can sometimes be traced to similar historical, cultural or intellectual origins. We try, in other words, to identify and communicate the general form of the forms of thought.

The particular course discussed here had two related objectives: to teach undergraduate students how to use the World Wide Web (as well as other Internet resources) and to help them to conduct research that transcended the boundaries of traditional academic disciplines while they learned computer-mediated research methods and critiqued ways of knowing. In "Convergent Computer Research," in short, we studied both disciplines and methods, including learning methods, by using many resources. Four print texts were included: a novel by a National Book Award winner that deals in part with the effects of technology on contemporary civilization; an informational guide to using the Internet; a book-length critical essay about computer uses and abuses in our educational system and wider society; and a philosophical inquiry into the nature of symbolic thought in human beings. These texts were selected because each is written in a different genre (fiction, informational, scientific and philosophical exposition) while touching upon many common threads: the roles of technology in our lives; the differences between signs and symbols; the connections between ideas and information; ways of thinking in the disciplines.

These print texts as well as a number of hypertextual and electronic texts were essential, leading both students and teacher to explore the six guiding course topics: (1) What is the difference between information and knowledge? (2) According to the authors we read and interpreted, what constitutes "knowledge" in the sciences? In the humanities? (3) What is "convergent" knowledge? (4) How do human beings think, know and feel? Does the computer affect human cognition and emotion? If so, in what ways? (5) What role can, should, or might computers play in the making and dissemination of knowledge? and (6) How can scholars in your field use computers most effectively for research?

As they worked with the intellectual substance of "Convergent Computer Research," students were taught to use World Wide Web search engines, directories and indices to conduct both individual and collaborative research. They worked in campus labs fully equipped with state-of-the-art web browsers and Internet capabilities as well as with various hypertext programs and a number of writing and collaborative programs. At the same time students beta tested the prototype for a new SFSU conference system, Conferencing on the Web (COW), a Web environment that provided them with a series of topics (those outlined above), the choice of joining conversations or starting new ones, and the opportunity to share their works-in-progress with one another while also critiquing and debugging COW, the invention of SFSU's Computer Science Chair, Professor Gerald S. Eisman, and a cadre of dedicated graduate students.

Encouraged to collaborate with one another as well as to experiment with various ways of presenting their knowledge--by learning hypertext markup language (HTML), by developing home pages, by integrating graphics, sound and text--students learned to see "knowledge" as a social construct that cannot be restricted to individual disciplines. They began to understand that "knowledge" changes with time, information, thought, synthetic and analytical expertise. Participants also learned the differences between sign and symbol, between scholarly technique and scholarly purposes, between gathering information and synthesizing such information into convergent knowledge bridging various disciplines. Working collaboratively with peers, with expert student volunteers who had previously taken the course and with the instructor, students began to distinguish new ways of knowing and new ways of making meaning in their own and other disciplines. They also grappled with issues of censorship, new forms of literacy, the purposes of education, the positive and negative effects of classifying knowledge by using traditional disciplinary definitions.

Preserved on the COW, much of the participants' lively on-line discussion showed two tendencies: first, a marked movement towards more expansive definitions of knowledge and wisdom; and secondly, growing skill at making meaning through the use of computer-mediated tools as well as print texts and collaboration. These transcripts of electronic talk demonstrated to me that NEXA students were genuinely interested in discovering the assumptions beneath our everyday concepts. As they did so, they learned to use new instructional technologies, the World Wide Web in particular, to find information, to decide on the value of this information, to synthesize facts and data into knowledge and ideas. Most students were not only eager to explore ideas and beliefs they had once taken for granted but delighted to join a collaborative learning environment, one in which the process of discovery among peers was considered as valuable (or even more valuable than) a teacher's lecture.

Participants' final projects took many shapes. Many constructed Web sites--exploring the meaning of death from various perspectives; compiling holistic health resources, organizations and definitions; contemplating the nature of consciousness; creating computer art projects that transcended the boundaries of painting, drawing, film and writing. Others developed hypertext documents that combined image, language and sometimes sound, focusing on areas as diverse as artificial life and the meaning of artifacts and architectural finds at Knossos. Still others wrote essays, developed MOO sites, created multimedia presentations. These final projects, which were presented to the class at the end of the semester, were in one way the culmination of the course, showcasing as they did the product of many months of thought, collaboration, reading, writing and learning. But as impressive as these final products were, they marked the beginning rather than the end of scholarly quests for these young scholars. Many have gone on to discover new ways to explore convergence using the World Wide Web, new methods to alchemize research into powerful ideas, new paths which they now explore at the graduate level and beyond the boundaries of school.

"Convergent Computer Research" has opened many new vistas for me as well as my students, helping this English professor to understand not only how to provide my students with an authentic education--one based in part on the principles of critical pedagogy, in part shaped by feminist scholars and others--but also how to develop new forms of teaching and learning appropriate in the remarkable medium of the World Wide Web. While new educational possibilities made possible by the Web are transforming literacy, too often teachers are the stragglers instead of pioneers. A road map would have been convenient but impossible to provide since this terrain has yet to be mapped. But we can and should join our students to draw the map ourselves.

Kids Interacting With Kids World-wide Through Web-based Visualization Technologies: *The Kids as Global Scientists Project*

Nancy Butler Songer
School of Education and Institute of Cognitive Science
University of Colorado
Campus Box 249
Boulder, CO. 80309 USA
songer@stripe.colorado.edu
http://stripe.colorado.edu/~kgshtml/Home.html

The World Wide Web offers dramatic new teaching and learning opportunities for K-12 education. With extensive funding from the National Science Foundation Application of Advanced Technologies and Presidential Faculty Fellows programs, the *Kids as Global Scientists Project* (KGS) began its fourth year of exploring the learning potential of emerging Web-affiliated tools, including professional conferencing software, Java Applets for real-time weather imagery, and on-line classroom activities which are continuously updated with new science information.

Over the past four years, approximately 4500 pre-college students in sixty world-distributed locations have learned to use the Internet telecommunications network, Netscape-based curricula, and real-time visualization software resources for learning about real-time events in environmental education. In this time period, KGS classrooms have included sites as diverse as: Albuquerque, New Mexico, USA; Galena, Alaska, USA; Harlem, New York, USA; Adelaide, Australia; Alberta, Canada; San Juan, Puerto, Rico; Shatin, Hong Kong; Zikhron-Yaakov, Israel; Aberdeen, Scotland; and Rovaniemi, Finland.

Student work is guided by an eight-week curriculum, called *Kids as Global Scientists,* developed by Dr. Nancy Butler Songer and colleagues at the University of Colorado, Boulder, CO. USA [http://stripe.colorado.edu/~kgshtml/Home.html]. In addition, students use their own data collection devices and state-of-the-art visualization software, including *Blue-Skies* developed by Dr. Perry Samson and colleagues at the University of Michigan, Ann Arbor, MI. USA [http://groundhog.sprl.umich.edu], to collect, share, and analyze real and near-time data. In many cases students learn to use interpret and use powerful imagery, including infrared and visible satellite imagery, temperature and wind vector maps, weather animations displaying a series of images, and many others. Then students use the imagery and information as they perform web-based activities that encourage them to predict the path of a recent hurricane and use current news reports and personal stories from individuals experiencing the storm to interpret the degree of damage or to develop scientific explanations of the occurrence. In another example, students investigating a storm originating in the Pacific Ocean near Hawaii downloaded current weather imagery of the event and corresponded with students nearby as a means to learn about the stormís patterns and characteristics. As the storm moved eastward, new studentsí expertise was called into play. In this and other ways, students learn to capitalize on the best features that web-based learning activities can provide, including real-time information and imagery, real-time conferencing with groups of distributed individuals, and other opportunities to interact with the information and individuals the telecommunications tools provide.

The result of the development and orchestration of projects like this is a unique and highly motivational educational experience which allows pre-college students an opportunity to evaluate, discuss, and analyze dynamic and currently-occurring science. Many students comment that they greatly appreciate this opportunity to get a rich and more meaningful view of science, where personal communication is used as an important vehicle for collaboration and learning, rather than utilizing a majority of static or impersonal sources for information. This presentation will feature a variety of multimedia illustrations of the successes and continued challenges of the past four years of the National Science Foundation-funded *Kids as Global Scientists* project.

Project PILOT - The Application of Telematics in Co-Operative Education

Andy Stone
Institute of Educational Technology
Open University
UK
Andy.Stone@open.ac.uk

Introduction

Project PILOT [PILOT 1996] is a joint project with two other universities which will deliver a prototype of a support system and matching service for all parties involved in 'co-operative education' (known as 'student projects' and 'placements' in the UK).

This support system includes a number of strands which meet the needs of the distinct user groups identified (which are students, placement units, and businesses). The main strand is a clearing house system, using a distributed database of candidates and vacancies for work placements. The database engine has been developed in MS Access.

For students, other strands of the PILOT system include a repository of support materials for reference (i.e. delivering "learning to learn" through telematics, elaborated upon in [Karran & Lefrere 1996]), plus links to more general support resources on the web. An example is the web-based guides to British towns and cities [Knowhere 1996], which are useful since many placements involve working in a different town to the one in which a student is at college.

A "best practice" repository system will be set up for placement units and companies, with copies of and/or pointers to case studies and other material that is relevant to placements. This will be supported by email-based discussion groups which will facilitate communication within the group and a source of feedback for the PILOT development team.

The features of this system were identified and are being refined through surveys and testing with placement officers and tutors, as well as managers and mentors in industry and commerce. As mentioned above, this process of refinement will be ongoing, via the email groups. This ongoing research will help to achieve a continual state of competitive advantage for all players, by realising Nouwens and Bouwman's [Nouwens & Bouwman 1994] second-order effects (i.e. innovative uses of telematics) *in order to* facilitate their first-order benefits of increased effectiveness and efficiency.

For example, the manner in which PILOT integrates with the existing placement unit function had to reflect the heterogeneous nature of the placement system; through the interviews, our research found that there is no national standardisation of this process. Complementing this, support for the formation of a discussion group to address this issue and to disseminate best practice was also identified in these interviews. Within the same interviews, methods to facilitate improving inefficiencies were being requested for inclusion in a telematics integrated application.

PILOT'S User Base & Accessing PILOT

Students and placement units will be accessing PILOT from academic sites, traditionally with high bandwidth (2-140 Megabit/sec on JANET/SUPERJANET respectively [JANET 1995]); on the other hand, businesses may have anything from a dial-up account (up to 28.8kbps) to a leased line (ISDN-T3), and also might not want a full web browser on a staff members' desktop (in some interviews, we discovered that in some companies, running a web browser from a desktop can be a dismissable offence).

In order to cater for the this latter group, a self-contained 'PILOT browser' is being developed. This is being authored in IconAuthor 7.0, which supports HTML as a data type, as well as supporting database connectivity. This will facilitate internet access to the PILOT site only, and also allows scope for releasing a CD of the PILOT database, probably on a subscription basis).

Other interfaces include fax-on-demand, but will not be implemented in the prototype stage due to cost constraints. Similarly, PILOT is being designed so that interactive telephony (through landlines and digital cellular networks, including the exploitation of Short Message Service technology on the latter) can be used as a conduit.

The use of audio-based (as opposed to text-based) resources is being examined as a complementary medium within the PILOT system. For example, as well as having textual case studies for the three groups to examine, there will also be documentary-style audio archive material, probably available as RealAudio files as well as stored for accessing through conventional telephony interfaces.

Conclusions

At the time of writing (August 1996), Project PILOT is still in its initial stage of funding and development (BT's funding is until December 1996). However, the web site [PILOT 1996] is up and running, and has a "visitors book" facility which also has the option of registering on our mailing list (email or fax, former is preferable).

In addition to the plans outlined in this paper, there is also the potential for expanding PILOT to other levels of education (i.e. for placements outside postsecondary education), and perhaps to other sectors where skills matching is needed. It is also intended to explore the development of a more international service (i.e. Europe and worldwide).

Apart from expanding the service, there is considerable scope for further research within the PILOT facility as it stands: desired outcomes include the ability to follow user behaviour and interaction in (and outside of) the main PILOT web site; i.e. which links are followed the most, possibly suggesting areas in which PILOT's own core facilities can be improved. An prototype of how this could be achieved is being developed at the Open University [Stratfold 1996].

References

[PILOT 1996]. http://mcie002.open.ac.uk/PILOT/

[JANET 1995]. http://www.ja.net/superJANET/superJANET.htm

[Karran & Lefrere 1996]. Karran, T.J. & Lefrere, P. (1996). Using Telematics To Support Personal Development in Projects and Work Placements. Paper delivered at ëTechnology Transfer and Innovation 96í conference, 1-3 July 1996, London.

[Nouwens & Bouwman, 1995] Nouwens, J. & Bouwman, H. (1995). Living Apart Together in Electronic Commerce: The Use of Information and Communication Technology to Create Network Organisations. http://shum.huji.ac.il/jcmc/vol1/issue3/nouwens.html

[Knowhere 1996. http://www.state51.co.uk/knowhere/index.htm

[Stratfold 1996]. Stratfold, M.P. (1996) "Tracker", http://trout.open.ac.uk/tracker.html

Acknowledgements

The PILOT project team: Stephen Brown, Patrick Haverty (De Montfort University), Paul Lefrere, Andy

Stone (Open University), Christine Doherty, Terence Karran (University of Lincolnshire and Humberside). Project PILOT is funded by British Telecom.

Special thanks to Matthew Stratfold for assistance with preparing this paper.

Use of the Web as a Tool for Interactive Learning

Eugene S. Takle
Department of Agronomy and Department of Geological and Atmospheric Sciences
Iowa State University, Ames, IA 50011 USA (gstakle@iastate.edu)

Michael R. Taber
Department of Agronomy and Department of Geological and Atmospheric Sciences
Iowa State University, Ames, IA 50011 USA (mtaber@iastate.edu)

1. Introduction

We are developing an Web-based course on global environmental change with goals of both scalability and asynchronous delivery, while maintaining a level of personal contact typical of a class of 50 students. We assert that the Internet offers unique possibilities for interactivity and scalability. The course (Takle & Taber, 1996) has evolved from a conventional lecture course with hand-outs and writing assignments to one in which we have attempted to preserve desirable features of the conventional course and add Web-based capabilities. The course homepage (URL: http://www.physics.iastate.edu/gcp/gcp.html) contains a syllabus summary with links to course information and grading, schedule of lectures, list of students and their e-mail addresses, required reading, reference lists, assignments, other related homepages, recent news releases, a forum for student-student and student-expert dialog, information on national and international global-change meetings, a forum for exchanging ideas about teaching on the Web, and a panic button for navigational help. The course addresses physical, biological, economic, social, and political implications of global change. Key objectives of the course are (1) to demonstrate the interconnectedness of components of the earth system, (2) to instill in students the value of authoritative peer-reviewed literature on global-change issues, and (3) to engage students in dialog among themselves and with outside experts on global-change topics.

2. Information Retrieval

The course uses the Internet to link to time-sensitive databases such as current global weather patterns, sea-surface temperatures, stratospheric ozone measurements, and vegetation indices. Other links include global-change conference information, modules assembled by government laboratories, and other information normally found in textbooks for which current data are available on the Web. Students tend to learn concepts better if use is made of current data, so we have tried where possible to use current data in interactive activities. Transcripts of the lectures, (including visuals) are posted for review by students. In a future expansion, we intend to supplement the text and images with audio. Students can search the Parks Library at Iowa State University and other on-line library resources directly from the homepage.

3. Electronic Dialog

An electronic dialog directly accessible from the homepage enables students and others (e.g., alumni, outside experts) to contribute to the course database. Students enter assignments, questions, essays, literature searches, arguments, or other information into databases organized by the instructor. The electronic dialog (1) extends class discussion beyond the lecture period, (2) encourages student interaction with the instructor outside of class, (3) allows students to dialog with outside experts on a particular issue, (4) allows the instructor to organize and monitor student interactive exercises, and (5) helps students to find new information (on the Web or otherwise) suitable for inclusion on the homepage. Each lecture has an electronic dialog page that allows students of pose questions or report information they have read relating to the topic. Responses may come from instructor, other students, or outside experts contacted specifically to address that issue. It also allows students to interact with counterparts at other universities and countries.

Some lectures have pre-lecture assignments on the Web to help students build an experiential base for the lecture material. Access to pre-lecture activities is closed at the time lecture begins as an incentive for students to prepare before class. Post-lecture Web activities lead the student to apply concepts from lecture to new situations or connect to future lecture topics. Students are required to summarize papers from the peer-reviewed literature and to write essays on how global change relates to their particular disciplines. These reports, by being generally available over the Internet, become part of the database for the course for access by future students. Having students summarize current research articles is a very efficient way for an instructor to keep current in a rapidly changing field that spans many disciplines.

Ethical issues relating to global change are posed by the instructor, and students are asked to respond to the stated issue or to responses by other students. This has been one of the popular elements of the course and the major source of electronic interaction among students. We also established an in-class audio/video remote dialog over the Internet with an internationally known author who discussed his recent book (excerpts of which had been read by students in advance) with students for about half an hour.

4. Student Reactions

Student response to the course, as determined from extensive surveys, was generally favorable, especially to the electronic dialog on ethical issues and audio/video dialog with an international expert. Several impediments were cited, most notably the data transfer time over the Internet. Students also disliked reading more than about 1-2 pages of text off the computer screen, and preparing for exams was more difficult. On the other hand, 84% of students likely would connect to the homepage at least occasionally after graduation from the University. This offers continuing education opportunities such as retaining connectivity with graduates who become K-12 teachers.

5. Advantages and Disadvantages of Web-Based Courses

There are both technical and pedagogical advantages to having a Web platform for a course. Technical advantages include: global accessibility, 24-hour availability, low cost, high scalability, asynchronous delivery, and ease of updating materials. Pedagogical advantages include the following: students can learn at their own pace, students can engage in remedial opportunities (relaxes course prerequisites), students have access to enriching opportunities, laboratory/group discussion can be included in any course, students are offered practice in rapid retrieval/synthesis of information, public writing and international perspectives are promoted, cooperative learning is easily incorporated, interactive exercises can be implemented with current data, student-to-student dialog on course and ethical issues is encouraged, and electronic publishing is available.

Potential disadvantages of this technology include the following: students don't like to read large volumes of text off computer screens, technical interruptions, time delays in bringing up images, and the possibility of sabotage or disruptive writing. These have not been problems for our course to this point.

6. References

Takle, E. S., and Taber, M. R. (1996). Use of the Internet as a platform for a university course on global change. Preprints, Fifth Symposium on Education, Atlanta. American Meteorological Society, J33-J34.

Acknowledgments

Doug Fils of the International Institute of Theoretical and Applied Physics has provided invaluable assistance in establishing the advanced functions of the homepage, particularly the electronic dialog. Funds for expansion of this course were supplied by the Office of the Provost, Iowa State University.

The Development of Multimedia Courseware of Japanese Culture via Internet

http://www.art.uiuc.edu/tea/

Rumi Tatsuta
Dokkyo University
tatsuta@dokkyo.ac.jp

Kimiko Gunji
University of Illinois
kimiko@uiuc.edu

Mary Ellen Michael
University of Illinois
mmichael@uiuc.edu

Introduction

Many US universities have Japanese Language classes but a few universities have Japanese culture class.

Japanese Tea Ceremony has been taught as a credit course at School of Art and Design, University of Illinois at Urbana-Champaign since 1965. The enrollment number is limited to 30 students because of the limitations of facilities and staff. However, the number of students who wish to enroll has increased in the last 10years; and every pre-registration, over 100 students have been requesting to enroll in this class. In order to meet overwhelming requests from students, Professor Gunji who is professor of School of Art and Design , turned to computer-based instruction as a solution. In January 1995, Professor Gunji completed the text of Japanese Tea Ceremony, collaborating with Tatsuta who was visiting professor of University of Illinois and Michael who is in the Educational Technologies Assistance Group(ETAG), Office of Instructional Resources(OIR). We have used this courseware in two classes; (1)East Asian Languages and Culture 150: Introduction to Japanese Culture (2) Art & Design 209: Tea Ceremony and Zen Aethetics.

Evaluation from students

We have created following questionnaire and we got 30 answers; 13students of Art and Design 209, 17 students of East Asian Language and Culture 150.

1. What aspects of the tea ceremony did you learn form the web site?
2. List seven or more Japanese customs and special cultural aspects that you have learned from this web site.
3. List strong aspects of this format for the introduction of tea ceremony.
4. Your advice for the improvement of this site will be greatly appreciated.

Phase II project:

1. Exploring and re-designing the interfaces of this courseware
2. Adding another chapter on Tea Aesthetics - Beauty in Tea: Wabe-Sabi
3. Translating the text into Japanese

References

Tatsuta, R. Gunji, K. & Michael, M. E.: Development of Networked Courseware at University of Illinois and Some Issues. T. of Japanese Society for Information and Systems in Education, Vol.95 No.4(1995)

Sensational Student Science Simulations: Physical Prestidiations done in HTML

Bonnie B. Thurber
Baker Demonstration School
National-Louis University
665 Brierhill Road
Deerfield, IL 60015-4405
USA
bthu@nlu.nl.edu
http://www.iceberg.org/~bthu

Vito M. Dipinto
National-Louis University
2840 Sheridan Road
Evanston, IL 60201
USA
vdip@nlu.nl.edu

The nature of science is both empirically and socially constructed. The former is daunting enough to teach and learn for middle school students. The latter seems to be beyond the do-able at this level. However, we have been engaged in a study over the past four years exploring how both of these aspects of science can be taught and learned by middle school students. The key element is the use of technology, specifically authoring in HyperCard or hypertext markup language (HTML).

Under the title Sensational Student Science Simulations we have had eighth grade students at the Baker Demonstration School of National-Louis University designing, implementing, and publishing chemistry and physics experiments. Students choose an interesting question or topic in chemistry or physics to investigate.

The socially constructed scientific knowledge is acquired by preparing a multimedia document either in HyperCard or HTML. The students provide a simulation of the actual investigation, include all the appropriate background scientific information, and introduce the essential scientific process and concept that this investigation demonstrates. The critical element in the learning is the sense of audience: a group of 4/5th graders at the Demonstration School and users of the Internet. In the community of scientists, the audience is those involved in similar areas of research. In the community of middle school students, the audience is other students in the school and people on the internet. This project helps instill in the students a sense of the possibilities inherent in science and technology and helps to develop the skills and interest in learning they will need to continue through life as seekers of knowledge.

This process, by including both the empirical and the social aspects of the physical sciences as part of the teaching/learning experience, provides the students with the opportunity to recreate scientific studies as well as to act as teachers in turn. This combination encourages both exploration for the sake of knowledge and the sharing of that knowledge with others. It simulates the research/publication/research cycle followed by practitioners. The inclusion of the HTML and Internet pieces in the project expand that experience by allowing the students to take their learning to the world and to serve as instructors or guides to a larger audience. This project helps instill in the students a sense of the possibilities inherent in science and technology and helps to develop the skills and interest in learning they will need to continue through life as seekers of knowledge.

MENTOR: Internet Search Advisor and Information Retrieval System

A. A. Tsinakos and K. G. Margaritis
Department of Informatics
University of Macedonia
54006 Thessaloniki
GREECE
Tel : +30-31- 891 891
E-mail: tsinakos/kmarg@kirki.it.uom.gr

Introduction

Internet provides access to hundreds of gigabytes of information through a variety of wide-area filing, information retrieval, publishing and library access systems. This rapidly growing data volume and diversity in Internet has created significant problems related to the efficiency and accuracy of the information retrieval. To make effective use of this wealth of information, user needs means to locate information. Additionally, information in existing Internet repositories is heterogeneous, inconsistent and sometimes incomplete [Bowman and Danzig and Manber and Schwartz 1994]. This fact increases the difficulty of the above mentioned problem. In the past few years, a number of such resource discovery tools have been created such as: 1) Internet Browsing and Exploring systems, such as Gopher, Hytelnet, Global Network Navigator, 2) Subject - Oriented Search systems, such as WWW, Virtual Library, Yahoo, USENET Frequently Asked Questions Archive, 3) Word - Oriented Search systems, such as Lycos, Web Crawler, Knowbot, Archie, WAIS and have gained wide popular acceptance in the Internet.

Further models originally developed for Artificial Intelligence research, have been applied to Information Retrieval leading to the development and evaluation of intelligent retrieval models for text documents, such as those found in bibliographic databases. These retrieval models specify strategies for evaluating documents with respect to a given query, typically resulting in a ranked output. Hypertext researchers, on the other hand have emphasized flexible organizations of multimedia "nodes" through connection made with user-specified links and interfaces that facilitate browsing in this network of links. A number of approaches to the integration of query-based retrieval strategies and browsing in hypertext networks have been proposed. The I3R system [Croft and Thompson 1987] and the medical handbook system described by Frisse, for example, use query based retrieval strategies to form a ranked list of candidate "starting points" for hypertext browsing.

Finally a number of probabilistic retrieval models for hypertext have been proposed [Frisse and Cousins 1989] [Savoy and Desbois, 1991]. These models view hypertext links as specifying important dependencies between hypertext nodes. The aim of the retrieval strategies based on these models are to improve the effectiveness of retrieval and to provide better starting points for browsing [Croft and Turtle, 1993].

In this paper we present a propose-and-revise system which automates the construction of a search strategy (in a specific domain) for Internet based information retrieval, in order to help novice or non novice Internet users to access and retrieve information using a variety of Internet search engines and information resources.

Overview description of MENTOR system

MENTOR system can be analysed in five levels 1) User level: The system can be accessed by one or multiple users and this is the starting point of user - system interaction. 2) Data Input-Output level: The front - end interface of the system, implemented in HTML, in which the user is allowed as first step, to insert his selections in a dialogue box and additionally to receive the results both of the proposed search strategy and the Internet search. 3) MENTOR's Advisor level: At this level the system implements and combines user inputs with the expert suggestions (using the pre-stored Librarian, Internet and Domain expert knowledge) in order to report the preferable search strategy, that is suggested to be followed, back to the user. A detailed description of this level is given in the following section. 4) Query Transformation level: This level is responsible for the transformation of the proposed search plan to individual queries towards Internet search engines. 5) Information Retrieval level: Here the system reaches the pre-selected search engines (from third level), and reports their results back to the user (second level) in HTML form.

Implementation Decisions

The implementation of MENTOR system can be separated in two stages: a) Implementation and development of the interactive component (levels 1,2,4 and 5), in HTML3 and CGI (Common Gateway Interface) scripts. b) Implementation of the Advisor component (level 3), in Common LISP.

MENTOR' s flowchart is as follows: STEP 1: User accesses MENTOR via WWW and inserts his inputs to the system (levels 1,2). STEP 2: User inputs are transferred to the Advisor Component so that the search strategy is determined, (level 3). STEP 3: The results of MENTOR' s Advisor Component are reported back to the user, (level 2). At this point, the user can modify the suggested search plan and proceed either to next step (information retrieval, levels 4,5) or to access again the Advisor Component (level 3), pursuing a new search plan. STEP 4: MENTOR reaches the selected search Internet engines (level 4) and reports their results (level 5) back to the user (level 2). The most important step, is step 2 where MENTOR' s Advisor Component is reached. Due to the complexity of this component a detailed description of its functionality is given in section.

Mentor's Advisor Component analysis

The Advisor Component of MENTOR system comprises of: 1. An automated Knowledge Acquisition component: This will be responsible for the knowledge elicitation from a domain expert and for the transformation of the acquired knowledge to a Knowledge Base System (KBS) as a side-effect of a man-machine dialogue. The stage of knowledge elicitation requires three different kinds of domain experts in order three different kinds of KBS to be constructed. 2. Three different Knowledge Base Systems (LKBS, IKBS, DKBS): The first type of knowledge base system, named LKBS (Librarian Knowledge Base System), is going to be constructed based on the acquired knowledge from a Librarian expert (a person specialised in subject or word-related search). This KBS will include the top level rules, tricks and tips that the expert usually follows, in order to locate the information that he is interested in. Similarly, the second KBS, named IKBS (Internet Knowledge Base System), will be based on the acquired knowledge from an Internet expert (a person specialised in Internet information location) and will include again the top level rules, tricks and tips that the expert usually follows in order to retrieve a specific information from Internet repositories. Both LKBS and IKBS will include rules and knowledge which will be domain independent and furthermore can be used and reused independently of the user defined search term. The third kind of knowledge base, named DKBS (Domain Knowledge Base System), will include information provided by the domain expert, in the sense of related concepts or synonyms to the user defined search term, in order the potentials of the users search to be enhanced. All the information stored in DKBS will be organised in distinct domain dependent sub-knowledge bases. Every time a new knowledge elicitation happens, the acquired domain specific knowledge will be added to the DKBS. Consequently the information stored in DKBS is not static but on the contrary is increased gradually each time a new knowledge elicitation happens. 3. A Knowledge Integrity tool: The Knowledge Integrity tool will be able to detect inconsistencies in the domain supplied by each domain expert and report them back to them so that they can be removed. 4. A forward chaining Inference Engine: This Inference Engine is going to be initialised using the user's inputs. Therefore based on these, and in combination with the information stored in the three pre-constructed knowledge bases LKBS, IKBS and DKBS, it will apply forward chaining to find all the rules and the related concepts that contribute in order a search plan to be reached.

Further Work

In the past several years, the number and variety of resources available on the Internet have increased dramatically. With this increase, many new systems have been developed that allow users to search for and access these resources. This paper outlines the overall description of a propose-and-revise system which can be used as an intelligent agent in the construction of search strategies for information retrieval in a specific domain applied on Internet information repositories. Further research topics include:- Method of knowledge acquisition to be followed.- Representation of acquired knowledge and concept piling.- Identification of Internet expert.- Training of the Internet novice users in information retrieval aspects, (justification why the proposed search is preferable).

References

[Bowman and Danzig and Manber and Schwartz 1994] Bowman,C. M., & Danzig, P.B., & Manber, U., & Schwartz, M.F., (1994). Scalable Internet Resource Discovery: Research Problems and Approaches. *Communication of ACM*, 37 (8), 98-107.

[Croft and Thompson 1987] Croft, W.B, & Thompson R.H, (1987). I3R: A new approach to the design of document retrieval systems. *Journal of the American Society for Information Science*, 38 (6), 389-404.

[Croft and Turtle, 1993] Croft, W.B, & Turtle, H. (1993). Retrieval Strategies for Hypertext. *Information Processing and Management*, 29 (3), 313-324.

[Frisse and Cousins 1989] Frisse, M.E., & Cousins, S.B (1989*). Information retrieval from hypertext: Update on the dynamic medical handbook project, 1989*, In Hypertext '89 Proceedings,.New York ACM Press. VA 199-212

[Savoy and Desbois, 1991] Savoy, J., & Desbois, D. (1991). *Bayesian inference networks in hypertext, 1991*, In Proceedings RIAO 3, Paris: CID., VA 662-681.

Feedback and Adaptive Interaction for WWW-based courses

Yury Tsybenko
Department of Dialogue and Tutoring Systems
Glushkov Institute of Cybernetics
40, Prospect Academica Glushkova
Kiev, 252 022 Ukraine
Email: yury@tel.dlab.kiev.ua

1. Introduction

In educational settings Web serves as organised recourse for a collection of structured learning materials with a set of hyperlinks (local and remote) in it. These features are educationally useful, but appear to be so routing now. It is necessary to do a research of use WWW beyond this standard features. During the last years WWW has been extensively studied to determine its learning-support possibilities beyond that of offering a home page. [Schneider and Block 1995] distinguished four main levels of WWW use in Education: (1) the Web as information tool; (2) distribution of learning materials; (3) collaboration tools; (4) interactive educational applications. This paper concerns levels 3 and 4 of the Web use in Education.

2. Two ways of interaction within the WWW-based course

Like it or not, but assessment of a student's knowledge and tutor's feedback are necessary elements of education and training. Students cannot learn much by only browsing a hypertext. Learning must be active. There are two different approaches in providing a feedback in distant education. The first approach is based on a role of a teacher. In some forms of distant education the teacher remains the central component in the course [B. Collis, 1995]. The difference from a face-to-face form is that the students are at a distance of a teacher. A teacher presents learning materials via Web and communicates with his students asynchronously via email, computer conference or by publishing teacher's instructions on the Web. Such an approach requires an additional tutors efforts on answering student's questions and evaluating student's work.

The another approach [E. Schwartz et al., 1996], [F Linn, et al., 1996] is integrating distant educational power of WWW with interactivity and intelligence of Intelligent Tutoring Systems (ITS). Traditional ITS technique aims to develop some ITS features that usually performed by human teacher in the classroom. This technique is based on a model of a student. Student model was developed to help intelligent tutor decide which exercise to give to a student, when to interrupt, what level of explanation to give, and so on. This technology is also may be useful for distance education, when teacher cannot interact with a student directly.

However, it is difficult to transfer directly the ITS technology in the Web. The ITS teaching strategies impose some constraints on student behaviour in learning, typically, the system leads a student through the learning materials. In contrast, a Web student can freely browse through the materials available, the system never knows the path by which the student has arrived in a hypermedia document.

3. Studying at the WWW learning environment.

The WWW-based course on basic Internet technologies consisting of 30 lessons is currently developed at Glushkov Institute of Cybernetics, Kiev, Ukraine. Each lesson contains on-line tests, exercises and other activities that form interactive learning environment of the lesson. The two different types of interaction are implemented for the course. The first is asynchronous interaction in which tutor communicate with students by email. This form of interaction is used for sending tutor's instructions, answering student's questions and monitoring collaborative work within a group of students. The second is adaptive on-line interaction provided by the server-side software. This form of interaction is used for evaluating the results of student's studying. The interactive forms allow to offer a variety of assessment methods such as true/false questions multiple-choice questions, short-answer and simple essay questions. The forms are spread through the course materials on the Web in order to supplement the main presentation chunks of the course with interaction facilities. The lesson assignment is a set of tests that relate to the lesson topics and should be answered simultaneously. Each lesson contains from 5 to 7 different tests.

The course learning materials consist of lecture notes divided into chapters and lessons. Each lesson is a hypertext containing interactive on-line forms such as tests and exercises. After browsing the lesson materials a student has to perform the lesson assignment. In a typical interactive session the computer will pose a problem and ask the student to fill in on-line form. After all forms have been completed the lesson assignment are submitted for evaluation. The student's answer is assessed by the server software and the student's scoring along with analysis of student's mistakes are returned to the student. To improve his or her current scoring a student is allowed to perform each assignment several times. This is also useful for a student's self assessment. Each testing attempt has been evaluated by CGI program, which processes the student's inputs. The CGI program also generates new version of the test for the next testing attempt.

The students are registered formally for the course. This provides some access constraints on the course materials and allows the system to follow the student's performance. The student's inputs are recorded by the server-side software. This helps a tutor to control the student's studying basing on past dialog history. Until a lesson assignment has not been completed with satisfactory scoring, it is no reason to assume that a student has learned the lesson materials.

4. Conclusions and future work

Current research demonstrates the usefulness of adaptive interaction technique for WWW-based learning. There are a lot of recently developed innovations such ATM multimedia, virtual reality, interactive video that will enhance the learning power of the Web. The intelligent interactive applications will only compliment the above innovations by performing a routine work on assessment of a student's knowledge and monitoring her or his learning. In addition intelligent features of the Web is enhanced factor for student's motivation and persistence.

In the proposed approach comprehensive tests are implemented to diagnose a student's knowledge and skills. The special database records information about the student's interaction with the system. This information will allow an intelligent tutor (student model) to infer current state of student knowledge and to provide adaptive interaction with the student. The most interesting future work is to develop a consistent model for a Web student. From technical point of view the ability of server side software to perform nontrivial computation is essential. Therefore the idea of incorporating advanced learning software as a part of server side software is natural.

References

[Collis, 1995] Collis, B. (Ed.) (1995). On-line and distance learning. URL http://www.to.utwente.nl/`campus.html.

[Schneider and Block,1995] Schneider, D., Block, K.,(1995) The World-Wide Web in Education. URL http://tecfa.unige.ch/ tecfa/tecfa-research/CMC/andrea95/andrea.html

[Lin et al., 1996] Lin, F., Danielson, R., Herrgot, S. (1996). Adaptive Interaction through WWW. Proc of Ed-TELECOM'96 173-178.

[Self, 1991] Self, J., (1991) Formal Approaches to Student Modelling. Tech. Report AI-59. Lancaster Univ., UK.

[Swarz et al., 1996] Schwarz, E., Brusilovsky, P., Weber, G., World-Wide Intelligent Textbooks. (1996). Proc of Ed-TELECOM'96 302-307.

[Wenger, 1987] Wenger, E., (1987) Artificial Intelligence and Tutoring Systems. Los Altos: Morgan-Kaufman.

"Silicon Soils": QT Surfs the 'Net

Max A.Turner
Department of Soil Science, Massey University
Palmerston North, New Zealand
Email: M.Turner@massey.ac.nz

James H. Cairns
Ag/Hort Multimedia Studio, Massey University
Palmerston North, New Zealand

Introduction

Increasing costs of education are a major current talking point, and concern, of students and staff at Massey University in New Zealand. Added to this is the need to commercialize existing courses to meet demands for more flexibility in time and place of study by existing and future students. To cater for the needs of those seeking off-campus tuition a distance (extramural) teaching program exists within Massey University, to deliver print-based course materials. Last year, nearly 300 tonnes of paper was consumed in reaching these students by mail.

Technological developments now exist to streamline this process and combine the advantages of the traditional face-to-face education, enjoyed by resident students, with the potential convenience and efficiency of off-Campus study. Massey University is in the early stages of experimenting with teaching via the World Wide Web /Internet and this paper outlines the approach adopted with an initial Soil Science course from within the Agricultural and Horticultural Sciences Faculty; the first to be offered this way in New Zealand. A major objective of this course was to employ a relatively low-cost approach to preparation and delivery, yet retain a professional appearance standard for the students and staff involved.

Design

The course aims to take advantage of the unique features available with WWW browsers and the Internet. The approach taken was to create a relatively simple, inexpensive, interface for the course, with a minimum of graphics and where possible, making use of low-cost freeware or shareware packages. A maximum depth of three screens from the main directory was imposed for each link in order to prevent students becoming lost in "cyberspace".

Construction of Course

The majority of the server functions are script-based (Perl-scripted cgi's). Each separate function has its own script and cohesion is achieved by interprocess communication. A Perl script is also used as a "chat" facility. Chat communication is effected using customizable screen refreshes optimized at around 4-10 seconds. Lecture content was prepared as Microsoft Word documents and converted to .pdf files (using the Adobe Acrobat Exchange plug-in). These course materials are then displayed to the reader using the freeware Acrobat Reader viewer, for which the Netscape browser supplied has been already configured. Graphics were created using Corel Draw 5, Paint Shop Pro, and Gif Construction Set. Reading materials were scanned using a Scanmaker E6 flat-bed scanner and ultimately converted into .pdf files. All reading materials, help files, and installation routines are provided to students on a CD-ROM, mastered using a Plasmon RF4100 CD-ROM recorder.

When installed at the user end, students enter via a local home page from which they are able to access the reading materials on their CD-ROM. This is made possible since the drive identifier for the CD-ROM was inserted into the client's home page as a hidden argument, at the time of installation of the package. Access to the Web course server is achieved after connection to the Internet. Security is achieved by password protection using standard .htaccess files with passwords supplied by each student.

Equipment

The course runs on a Unix (Linux) platform using the freeware Apache server. It is housed on a Pentium 100 Intel PC with a 1Gb hard drive.

Use of Internet

To be eligible to take the course students must either have access to the Internet from a local provider, or a modem to connect to the University Intranet. We normally recommend the direct Internet connection route which saves on toll charges, except when a student resides close to the University, within the free-call telephone zone.

Students connect each week to the Silicon Soils server to obtain lecture material, news, and project information, to submit assignments and laboratory reports to the instructors, or to communicate through the "asynchronous" chat facility. The course coordinators are also available at set times during the week to handle, synchronously, queries about the course via the same chat facility. If students have queries outside these times, a message is left in the appropriate chat, for later attention. All correspondence in chat is via text entry only. Students also have the opportunity to correspond by email which acts as a backup facility and an alternative means of communication for them, especially in the early stages of the course, until such time as they have negotiated the installation phase.

Assessment

At this early stage, course evaluation is by the usual extra-mural system which involves an invigilated examination at a recognized, appropriate venue. Assignments and exercises are submitted via ftp from within the Netscape environment.

Conclusions

An introductory course in Soil Science has been developed for extra-mural, distance education study at Massey University for delivery via the Internet. Where possible freeware and shareware packages were used, including the server itself, to author an attractive, professional, yet economical set of scripted html documents.

Future Directions

The next step will be to create a front-end for the course to provide a user-friendly template for authoring other courses which follow this model. Also, assuming a favorable response from students, other "distance-less" courses are planned along these same lines including some with a more global perspective.

Acknowledgments

This course was prepared using a grant from the "Fund for Innovation and Excellence in Teaching" at Massey University. The most valuable assistance of the following staff at Massey University is acknowledged:

Dr M.J. Hedley, Ms Marion Trembath (Department of Soil Science) and Mr K.M. Harris (Ag/Hort Multimedia Studio)

The ETT Virtual Classroom: A Different Flavor

Jean Vermel
EFL Department, Informatics Department
Beit Berl Teachers' College
jean@beitberl.beitberl.ac.il

The project run by ETT (English as a Foreign Language Teacher Trainers) (*http://mofet.macam98.ac.il/~ett/vchome/vchome.htm*) connects three teacher training colleges in Israel in a joint venture, through collaborative learning, in creating a "virtual classroom" for the teaching of English as a foreign language using the Internet.

Our main aim was to use the Internet as a tool to locate resources, references, materials, and on-line sites that would be useful to a future EFL teacher. The Internet was used as the medium to fulfill the assignment given.

The participants in the project were students training to become English teachers (about 15 students from each college). These trainees worked in groups of six, two from each college, on one particular topic, for example

- using current affairs in the EFL classroom
- using extensive reading in the EFL classroom
- using music in the EFL classroom

The trainees collaborated via the Internet in order to create a portfolio of ideas and suggestions for their own personal use, and for their use as future teachers.

This final assignment consisted of three different parts:

- an individual section, which was made up of reflective questions dealing with the use of the Internet as a tool and working with "invisible partners"
- a pair assignment, done with their in-class partner(s), which required them to analyze one particular site in depth and show how the information in this site could be useful in their teaching
- a group assignment, carried out with all the members of the group from the three different institutions, which dealt with choosing eight sites they felt were beneficial to them as future EFL teachers, coming to an agreement about those eight sites, and writing short explanations of how those sites could be helpful.

While our journey on the Internet dealt with mainly sites related to EFL, our "hidden agenda" was to teach the teacher trainees the different programs used on the Internet. We worked with our respective trainees in a regular classroom lesson environment, and taught them use of electronic mail, "surfing and searching" on the Internet, discussion groups, IRC etc. This added a different flavor to our project as the trainees needed to participate in our lessons in order to learn the tools which would then allow them to complete their assignment using the ETT virtual classroom.

The project had its problems, of course, and they manifested themselves mainly in the technical area:

- mail not working properly and getting lost
- server down during class time
- the computer room not being open during the week so that the trainees could "practice," etc.

We also discovered that it was most important for us, the teacher trainers, to take a forward role in moderating our respective listservs. The teacher trainees, being unaccustomed to communicating in writing, were initially quite hesitant about sending short messages to their groups about any subject and there was a lack of true "communication" among the participants.

We believe that now that the trainees have experienced being part of a virtual classroom, they will not only use this media for themselves as teachers but will involve their future pupils in similar projects.

These and other lessons learned have formed the basis for our continuing work in the ETT Virtual Classroom and it is our hope to open the classroom to EFL trainees around the world in 1997.

References

Cummins, J. and Sayers, D. (1995). *Brave New Schools. New York:* St. Martin's Press.

Nunan, David. (1992). *Collaborative Language Learning and Teaching*. Great Britain: Cambridge University Press.

Riel, Margaret. (1992). "Cooperative Learning through Telecommunications."

The AT&T Learning Network. *Spring '92 ACM Conference on Computer Supported Language Learning,* Vol. 21 #3, ACM Press. Available URL:

http://www-cscl95.indiana.edu/cscl95/outlook/Outlook.14_Riel.html

Salomon, Gavriel. (1992). "What does the design of effective CSCL Require and How Do We Study Its Effects?" *Spring '92 ACM Conference on Computer Supported Language Learning,* Vol. 21 #3, ACM Press. Available URL:

http://www-cscl95.indiana.edu/cscl95/outlook/Outlook.62_Salomon.html

Turoff, Murray. (1995). "Designing a Virtual Classroom." Reprints of Invited Papers for 1995 International Conference on Computer Assisted Instruction ICCAI '95. Available URL: http://www.njit.edu/CCC/VC/Papers/Design.html

Acknowledgements

My two colleagues, Nili Mor of Levinsky Teachers' Training College (nili@mofet.macam98.ac.il) and Elaine Hoter of Talpiot Teachers' Training College (elaine@talpiyot.macam98.ac.il) played an equal role in creating and running the ETT Virtual Classroom, and we see ourselves as an inseparable collaborative team in all our efforts at the ETT site: http://mofet.macam98.ac.il/~ett.

'A Cantina i Bifarelli': a Java based environment for real-time multiplayer Card Games

Antonio VOLPENTESTA
GIUDA Lab.
University of Calabria, ITALY
volpente@giuda.deis.unical.it

Antonio GRECO
GIUDA Lab.
University of Calabria, ITALY
agreco@giuda.deis.unical.it

Vincenzo PISCULLI
GIUDA Lab.
University of Calabria, ITALY
vincenzo@sungiuda.deis.unical.it

Nicola FREGA
GIUDA Lab.
University of Calabria, ITALY
nfrega@giuda.deis.unical.it

Introduction.

In ancient days folks of Cosenza (a city of Calabria, Italy), used meet together at night in a tavern called 'A Cantina i Bifarelli. They played card games, as the Tressette, drinking wine and chatting about what happened in their city.

Today 'A Cantina i Bifarelli does not exist anymore, but we have realized a software system to renew its spirit.

At the present time, many Web sites give the opportunity to play card games and this testifies the interest for such a kind of applications in the Internet community (see Gamelan - *http://www.gamelan.com* or Jars - *http://www.jars.com*).

The system may be regarded as an environment that allows users to communicate each other by a chat service, while they are playing some card game on a virtual table. It has a client-server based architecture.

In the server-side we have a Java stand-alone application which manages client calls, message exchanges and status variables for any ongoing game.

The client-side is based on some applets for managing the end-user graphical interface and the communication with the server.

Our system is characterized by two main features:

- easy extensibility: a new card game may be immediately incorporated as soon as it has been implemented;
- finitely many card games can take place on different tables simultaneously.

In what follows we present the realized environment by briefly describing:

- the server architecture;
- the client/server communication protocols;
- the applets for the client architecture.

Moreover, we show the Web pages organization for an easy access to the environment. An extented version of this paper can be found at *http://sungiuda.deis.unical.it/java/HTML/paper.html*. In such a version we give some other details about the environment and we present rules to be known to play the TRESSETTE, an on-line game at 'A Cantina i Bifarelli.

Server architecture.

The server consists of a Java stand-alone multithreaded application which manages the communication with Java applets clients via sockets. It's architecture is outlined in [Fig. 1].

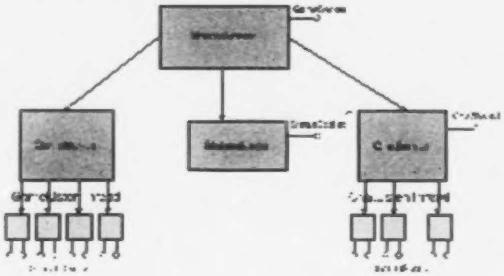


Figure 1: Server architecture.

The MasterServer is the main process; it starts offspring threads and manages the GameSocket.

The StatusServer handles the StatusSocket in order to satisfy a client request of server status parameters value.

The ChatServer handles the ChatSocket; it associates two sockets (Input and Output) and a thread to any connected client and it sends any received message to all clients.

The GameServer handles a single card game and it is started by the MasterServer as soon as a new player arrives into an empty room. A thread is associated to any player and it handles two sockets for the client-server communications.

Communication protocols.

We have considered three protocols:

StatusProtocol (SP), for the StatusServer and StatusClient communication;

ChatProtocol (CP), for the ChatServer and ChatClient communication;

GameProtocol (GP), for the GameServer and GameClient communication.

In the SP case, the client which is connected on the StatusSocket receives back a string message from the server of the following format:

STATUS: <Room1Status>|<Room2Status>|...|<RoomNStatus>.

Such a string is then interpreted and graphically displayed by the client for the end-user.

BEST COPY AVAILABLE

In the CP case, the client sends on the ChatInputSocket a string of the type:

USERNAME: <nickname>

where nickname is the name chosen by the end-user. The client sends messages on the same socket with the format:

PUBMESSAGE: <message>

and it receives from the server other client messages on the ChatOutputSocket with the format:

PUBMESSAGE: <nickname><message>

Therefore, the server simply collects the incoming messages, labels them with the appropriate nickname and transmits them to all clients.

The GP is actually a set of protocols each one of them is relevant to a particular card game. At the time being the only card game implemented in 'A Cantina i Bifarelli is TRESSETTE.

Applets.

We have an applet for each specific card game. Moreover, we have:

StatusClient: it allows to know which game has been selected in each room, which number of players is requested to start a game and who is waiting for playing around the table. Beside, it makes possible to select a room and to run the ZoomClient applet.

ZoomClient: it lets a user know the status of a selected room and choose a free position around the table. In case of empty room a user may even select a game for that room.

ChatClient: it accomplishes the user interface for the chat service. In the input side a user may enter his/her own nickname and messages; the output side is devoted to display messages of all users.

References.

'A Cantina i Bifarelli can be reached at *http://sungiuda.deis.unical.it/java/HTML/index.html*

An extended version of this paper can be found at *http://barabba.deis.unical.it/ Documents/WebNet96/bifarelli.html.*

Interactive Teaching Material for the World Wide Web: A Syntax Tutor Written in Java

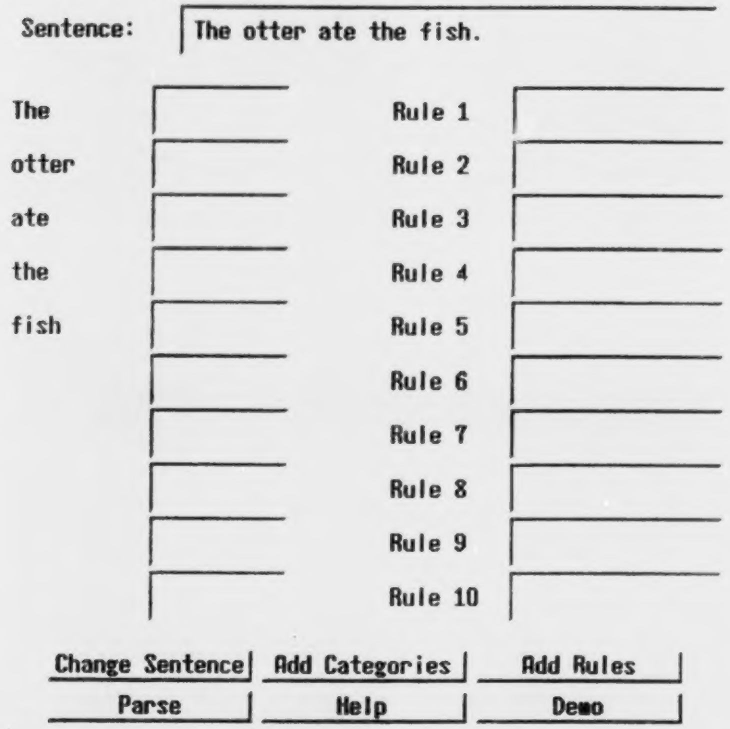
Abstract for WebNet-96, category "Educational Multimedia for the Web".

Ana von Klopp
Department of Linguistics
University of Wales, Bangor
Bangor LL57 2DG
Great Britain

email: avk@bangor.ac.uk
phone: +44 1248 38 22 68

This presentation describes an interactive tutoring system for teaching the basics of generative syntax to first year linguistics students. It was written by Ana von Klopp and Chris Dalton in Java, the C++-like programming language developed by Sun Microsystems which makes it possible to create interactive applications that can be accessed and run from anywhere on the Internet through a World Wide Web browser.

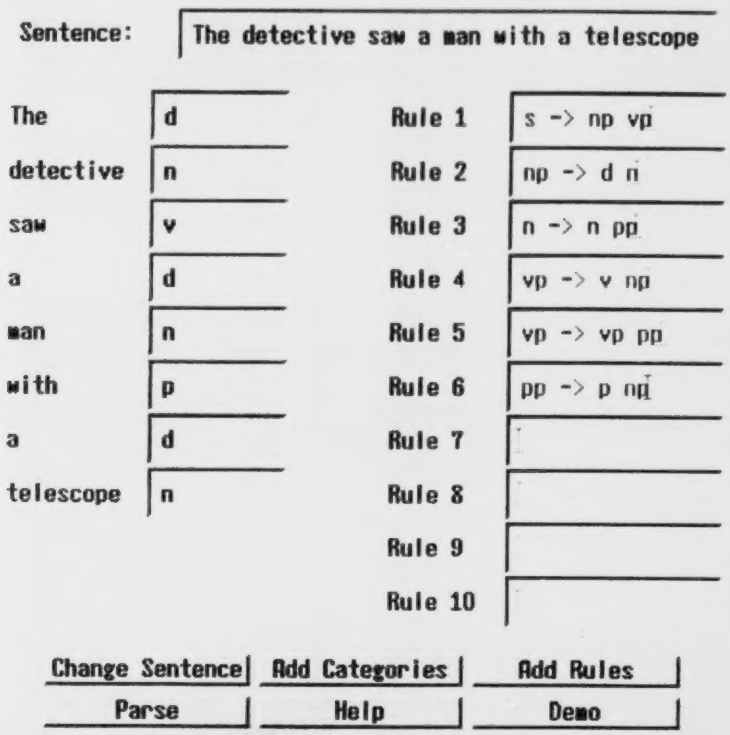
The **Syntax Tutor** allows the user to choose a sentence or phrase to analyse, assign syntactic categories to the words in the sentence and write a set of phrase structure rules. When the selections have been made, the program shows the parse tree(s) generated by the grammar or tells the user that no analysis can be found. The screenshots below show how the application works. The first image shows the appearance of the application on first visiting the page.


Sentence: The otter ate the fish.

The
otter
ate
the
fish
Rule 1
Rule 2
Rule 3
Rule 4
Rule 5
Rule 6
Rule 7
Rule 8
Rule 9
Rule 10
Change Sentence
Add Categories
Add Rules
Parse
Help
Demo

To select a new sentence, the user clicks on the **Change Sentence** button. This makes the sentence field editable, allowing the user to type in a sentence. Similarly, clicking on **Add Categories** and **Add Rules** make the category fields and rule fields editable. The next image shows the state of the application after the user has selected a new sentence, assigned categories to the words and typed in a set of rules.

BEST COPY AVAILABLE



Sentence: The detective saw a man with a telescope

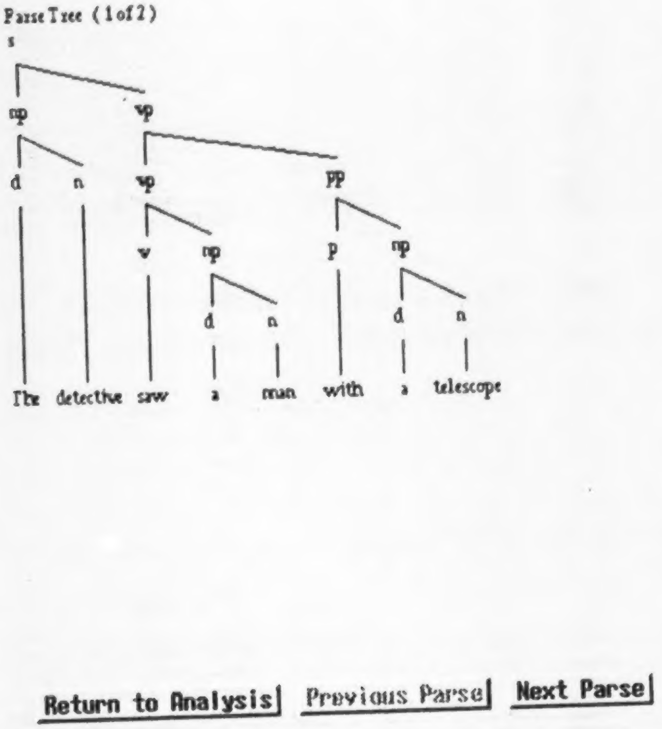
The	d
detective	n
saw	v
a	d
man	n
with	p
a	d
telescope	n

Rule 1 s -> np vp
Rule 2 np -> d n
Rule 3 n -> n pp
Rule 4 vp -> v np
Rule 5 vp -> vp pp
Rule 6 pp -> p np
Rule 7
Rule 8
Rule 9
Rule 10

The category labels for the words in the example above are *d* for *determiner*, *n* for *noun*, *v* for *verb* and *p* for *preposition*. The complex category labels in the rules, *s, np, vp* and *pp*, stand for *sentence*, *noun phrase*, *verb phrase* and *preposition phrase*. A rule such as *s -> np vp* states that a sentence can consist of a noun phrase and a verb phrase. (Standard abbreviations were used in this example, but any labels could be used as long as the user is consistent.)

When the selections have been made, the user clicks on **Parse**, and trees corresponding to the analyses generated by the grammar (if there are any) are shown. With the grammar given in the image above, there are two distinct structural analyses, one of which is shown in the picture below (the **Next Parse** and **Previous Parse** buttons are used to move between parse trees; **Return to Analysis** returns to the input display).

BEST COPY AVAILABLE



The applet is organised around two main displays which are constructed with components from the Java Abstract Window Toolkit: the input display accepts user input (sentences, category assignments and phrase structure rules) while the output display prints tree representations of successful analyses generated by the user's grammar (the input display is shown in the first two images; the output display in the third).

Sentences chosen by users are separated into tokens (words) next time they click on a button. The category input fields are adjusted accordingly, showing one labeled field for each word token. When the user clicks on the **Parse** button, the input format is verified before it is passed to a parser object as a `Sentence` object made up from `Word` objects consisting of a syntactic category and the graphemic representation, and a `RuleSet` object made up by `Rule` objects.

The parser object then runs its parsing method. A left-corner algorithm implemented as a while-loop rather than relying directly on recursion is used. This type of algorithm makes it easy to handle left-recursive rules and keep track of choice points (if one is found a copy of the current state are saved for each alternative which is then pursued in turn). As the parser proceeds, it constructs a bracketed string representation of the grammatical structure, and successful parses are saved to an array.

If one or more analyses are found, the applet shifts to the output display. Tree representations of the sentence structure are painted by sending messages containing a single string representation, the total number of successful analyses and the number of the current one to the display. The string is then parsed by a second top-down predictive parser, which generates a set of relative coordinates for the nodes of the tree which are linked and labeled.

The applet can be accessed at URL http://www.linguistics.bangor.ac.uk/java/lt/LingTutor.html.

BEST COPY AVAILABLE

Providing Enterprise Wide and Localized Searching at a Large, Decentralized Institution

David J. Wang, djw@gdb.org
George Ciervo, ciervo@gdb.org
Division of Biomedical Information Sciences
Johns Hopkins University School of Medicine,
Baltimore, MD 21205 USA
http://infonet.welch.jhu.edu/miss_smt.html

Introduction

Like many research institutions, the Johns Hopkins University is highly decentralized. This lack of centralization pervades the culture of Johns Hopkins and is evident in nearly every facet of the university's business. Not surprisingly, many of the university's WWW servers also lack a central point of management that facilitates effective access. As a result, it is nearly impossible to build a hierarchical navigation structure with links pointing to all the "most appropriate" locations. Even if such a structure were possible, the continued development of pages by divisions, departments, labs, offices, and other groups would render any such navigation arcane. In this environment, an institutional site wide search is imperative.

Furthermore, though many of these sites are small, some are reasonably large in scale. The inclusion of searching functions in these individual larger sites would provide an additional valid and often necessary navigation function. Unfortunately, many of these sites lack staff with the appropriate technical skills to implement such a search. While designing the Hopkins-Wide Search, it was decided that providing a mechanism to easily allow Hopkins' web administrators to make their site searchable would add substantial functionality.

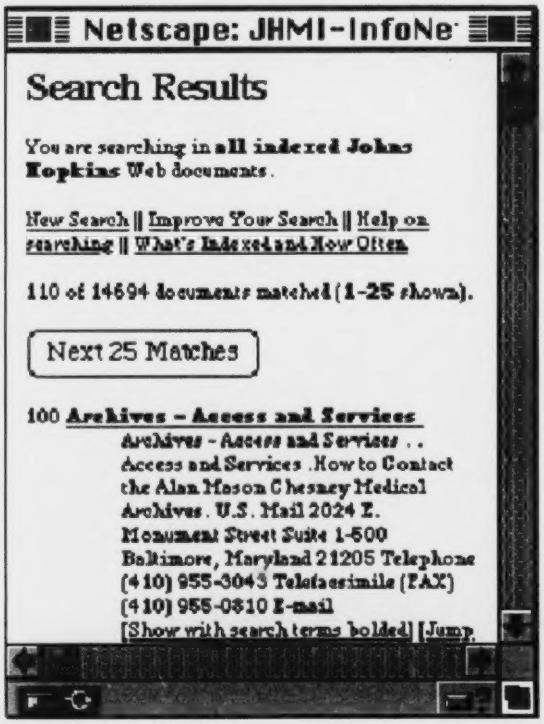
Providing a Uniform Search Interface for All Web Documents

After examining several possible searching solutions, Verity's Topic Internet Server was selected and installed. The server consists of a search engine, a remote indexer, and an end-user interface. The remote indexer is used to index multiple Johns Hopkins web sites. Once a site has been indexed, it is maintained as a separate "collection", allowing a user to search one or multiple collections. A form is provided [URL1] where webmasters can both register their site and obtain html-code for inclusion in their html pages. This request is forwarded to the InfoNet Development Group, and the site is included in the index within twenty-four hours.

Once indexed, the site is immediately searchable. The code provided to the webmaster is a form that makes searching that site possible. All the queries are against the index maintained by InfoNet. Thus, each site may be queried as part of the master index [URL2], which includes ALL the indexed sites, or as an individual collection [URL3]. Once the search has been executed, users are presented with a list of the top 25 results. They have the option of paging forward or backward through their results. Users also have the ability to improve their search by requiring that additional words be present, by weighting certain words, or by requiring the abscence of certain words. Finally, users searching in individual sites have the option to continue their localized site search or to expand their search to all the indexed documents.

Expanding the Hopkins Wide Search

In reviewing the server logs after the search had been put online, it was discovered that a number of queries were directory-type requests, i.e., users looking for personnel and contact information, and hence the need for a "one-stop-search" was recognized. Since there is no single and definitive online Hopkins directory, it was vital that this search not be limited to web documents, but also include other web-accessible information. As a result, the new search queries not only the full text web documents, but also the Medical Campus Phonebook/E-Mail directory [URL4], the Johns Hopkins Institutions Phone Directory [URL5] and the Johns Hopkins Expertise Database [URL6]. Frames are used to separate the results, where the number of frames is determined dynamically and is dependent on which searches return positive results [Figs. 1,2].



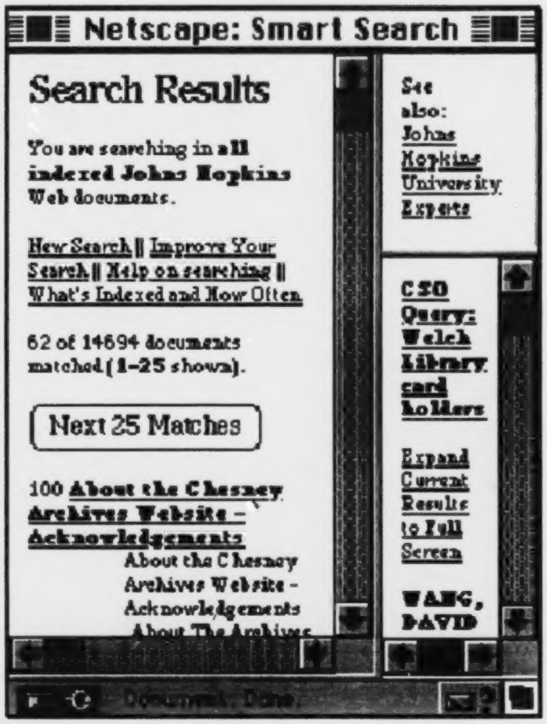
Search Results

You are searching in **all indexed Johns Hopkins** Web documents.

New Search || Improve Your Search || Help on searching || What's Indexed and How Often

110 of 14694 documents matched (**1-25** shown).

100 **Archives - Access and Services**
Archives - Access and Services . . Access and Services .How to Contact the Alan Mason Chesney Medical Archives. U.S. Mail 2024 E. Monument Street Suite 1-600 Baltimore, Maryland 21205 Telephone (410) 955-3043 Telefacsimile (FAX) (410) 955-0810 E-mail
[Show with search terms bolded] [Jump



Search Results

You are searching in **all indexed Johns Hopkins** Web documents.

New Search || Improve Your Search || Help on searching || What's Indexed and How Often

62 of 14694 documents matched (**1-25** shown).

100 **About the Chesney Archives Website - Acknowledgements**
About the Chesney Archives Website - Acknowledgements About The Archives

See also: Johns Hopkins University Experts

CSO Query: Welch Library card holders

Expand Current Results to Full Screen

WANG, DAVID

Figure 1: Single Result: *search on 'shuttle schedule'* **Figure 2**: Multiple Results: *search on 'david wang'*

Usage Analysis

There have been some encouraging trends in the usage of the search. Of 20,766 queries from May through June, 9,000 originated from other than the main index page. In addition, for the one week ending June 30, of 26,000 content accesses to InfoNet [URL7], approximately 1,100 were requests of the master search page.

Conclusion/Future Directions

Current plans are to incorporate the Personal Home Page Index [URL8], various searchable research directories (ex./ neurology, pathology), and other directory information [URL9] into this interface. In addition, as the number of accessible searches increases, it will be necessary to allow the user to customize which queries to execute.

For many people the web has moved beyond novelty and has become an important information gathering tool. This is true for many of the visitors of Hopkins websites, who are likely to be potential students researching their options for higher education, researchers looking for expertise or scientific information, individuals looking for referrals or health care advice, and internal students, faculty, or staff looking for intranet-type materials. As these types of requests continue to grow and as information providers realize this, we expect the number of web sites to continue to grow, further necessitating a comprehensive Hopkins-wide search. With our current hardware, software, and connectivity we expect

BEST COPY AVAILABLE

to be able to continue adding sites and indexes to our index for the foreseeable future.

URL References

[URL1-URL9] Please see: http://infonet.welch.jhu.edu/conferences/webnet96/urls.html

Acknowledgements

We would like to thank David T. Kingsbury and Jayne M. Campbell for their advice and support on this project.

Distant Education Materials via the Web - An Australian View

David Watson, Health Studies, University of New England, Armidale, NSW Australia
Email: dwatson@metz.une.edu.au

Kathy Watson, Education Studies, University of New England, Armidale, NSW Australia
Email: kwatson@metz.une.edu.au

Introduction

The Australian government decision to establish the electronic infrastructure, Education Network Australia (EdNA) and other related initiatives was intended to place information technology at the centre of educational reform yet "this infrastructure has made little impact on the organisation of teaching and learning services in the majority of universities" according to a recent report by the National Board of Employment Education and Training (NBEET. 1996) Higher education has been reprimanded for a lack of attention to community and industry needs by producing graduates lacking adequate exposure to, or awareness of, current technologies. In order to meet the community and industry demands for technology literate graduates, we are developing new modes of delivery to promote innovative ways to expose students to technology while creating the learning base required for understanding the subject material. For students isolated from traditional university campuses studying via distance education, the web offers a variety of telecommunication tools for course work delivery and feedback.

Universities are facing the challenge of reconciling the relationship between education, work and employment. The NBEET report specifically emphasises the importance of "information literacy" and suggests that education and professional development "should seek to link explicitly the three elements of this competence:

- use of information technologies,
- information research and presentation skills, and
- higher order systems thinking skills" (1996, p 94.)

Tasks given to students should be considered as tools to accomplish goals rather than tasks to be mastered. Use of the Web can provide students with many experiences: reading hypertext on the screen, using electronic mail, using navigational tools to find out more about areas of interest, and interpreting symbols, icons and command functions to access audiovisual resources. The motivation to learn blends with the tool use and does not need to be externally imposed as learning to use the 'tool' is instrumental in the context of the task.

The ability to access and utilise the information on the web for research purposes is consistent with the type of skills students will need in the world of work in order to problem solve in the world of work. Student exposure to web based technologies is seen by many potential employers as beneficial to the overall outcomes of a 'corporate plan' as organisations incorporate emerging global technologies. Leading education, business and industry organisations utilise the web for presentation of information via graphic images, hypertext links, and more recently audio and visual resources.

The benefits for the student/professional community from interactive web use are diverse. For the education sector, the ability to extend "classroom" teaching across a distance is a major benefit as staff resources are freed up to focus on other teaching duties. For students, benefits come from a larger resource network to draw upon and the chance to explore community values on topics beyond those available in their own communities. For isolated communities, classes available by live audio offer expanded opportunities not normally available due to location, student numbers, or suitable teaching staff. For professionals, the ability to tap into the latest research or to hear a refresher lecture on a topic can mean the difference between being up to date in their field or out of date with their profession.

Speaking, listening and learning...

"Recent advances in the compression of audio and visual signals has lead to a proliferation in use of live AV technologies on the Internet. Previously only accessible on mainframe frame computers using programs such as MBONE, audio was unavailable to most personal computer users connecting to the Internet" (Watson, D. M. 1995). The arrival of Real Audio and ability to provide "jukebox" selections of audio on demand, across computing platforms, has opened new doors of opportunity for distance education providers. The challenge for educators in the Internet community is how to best utilise this new opportunity.

"The wait for a suitable and affordable delivery system may be over as the Internet and recent advances in software (Real Audio - http://www.realaudio.com) now allow for the transfer of both graphic and audio materials in "real time". The ability to "listen" to live or prerecorded broadcasts while simultaneously reviewing graphic and or text materials is possible now. This new ability replaces the need to down load large files in order to listen to comments, medical related sounds (breath, heart, etc), speeches or lectures related to specific topics" (Watson, D. M. 1996).

The ramifications for advances in the transfer of information over the Internet is still being researched by many organisations. For health service education providers the Internet provides a rich resource of materials and views for use in teaching and research. The addition of *real time audio* adds new dimensions to opportunities available through networked systems and the Internet. For tertiary institutions teaching in health service related areas the addition of a functional audio delivery system via the Internet provides opportunities for students to listen to lectures or instructions within a suitable and flexible time frame. It provides a new medium for delivery of educational materials at a distance and allows for student review of previous discussions as needed.

Health service organisations inservice training programs that utilise audio presentations in combination with printed and graphic materials allow organisations to optimise the use of training and professional staff. The use of "pre-recorded" lectures and tutorials for access at a later date by shift workers, persons away on holidays during original sessions, those wanting to review the topics discussed, or by new employees adds to the value of an audio supplemented inservice program. Having "on hand" audio resources to describe a concept or function can be convenient and potentially life saving. In some cases, written instructions may not be understood where a verbal description or instruction may be more easily followed. Anyone who has tried to program a "machine" from a poorly written instruction book would confirm this point. If the machine in question happens to be a ventilator or other critical care device the ability to have instructions in an audio format may save a life.

In an age where more and more people are using electronic means to access information about problems related to their health needs, the ability to offer audio discussions to support printed and graphic materials is of key importance. Audio presentations can easily be tailored to meet language requirements presented by a multicultural, multilingual population or work force.

Bridging Language, Culture and Distance

UNE distance education students live all around the world. Working in International schools or other organisations in Malaysia, Thailand, South Pacific Islands and remote places in outback Australia, students are often seeking information about issues relevant to diverse multicultural societies. The Web provides students with a wide variety of resources and information related to interactive collaborative learning in text, pictures, graphics and sound. On the Web, students can find many examples of long distance learning designed to share knowledge, expand global understanding, facilitate cooperative problem solving by educators who use electronic field trips. The Global Schoolhouse International Project is one example, providing a model for 'active' learning and emphasising the need for students to develop research skills to cope with the vast amount of information.

Teacher education students are challenged to use a variety of technological tools to answer the question, "Why use technological applications to promote language and literacy skills?" in a unit designed specifically for students who study at a distance. Students were provided with the usual printed materials which included selected readings from sources such as: The Australian Dept. of Employment, Education and Training, International Journals, International Conference Proceedings, Professional Associations, UNE University Information Technology resources and Internet resources. The UNE information included the UNE WWW location, instructions for remote access to the UNE networked resources via a modem and Help Desk Information. The Internet resources included an annotated list of Educational Internet Sites worth visiting with URL locations such as: AskERIC Virtual Library, Children's Pages from Around the World, The Global Schoolhouse, and I*EARN (International Education and Resource Network). A hard copy of Frequently Asked Questions about Netscape Navigator provided an invaluable compilation of resources for the Web beginner.

Promoting skills needed for students to use information technology tools enables them to gain more control over their own learning. Therefore, the unit materials introduce some Internet resources to facilitate learning processes which also help students prepare their assignments. The assignments require students to: use the Internet as a source of information, increase their personal information technology skills, use criteria for software evaluation, and develop integrated curriculum frameworks which compliment language and literacy skills.

Ultimately, teachers must decide which educational tool is best for a particular task and if some form of technology can be effectively integrated into an existing curriculum. "These decisions are based on personal beliefs or views about learning processes. If teachers believe that all knowledge is constructed, instruction is a nurturing of these processes, and if they accept that literacy development is a social event, then their classrooms should provide experiences to facilitate this construction process" (Watson, K.K. 1993). Central to this view of constructivism is the notion of the learner as 'active' not just responding to stimuli, as in the behaviourist rubric, but engaging, grappling, and seeking to make sense of things (Perkins, D.N. 1991).

Preparing Today for Tomorrow's Opportunities

"Since Marshall McLuhan coined the term in the 1960's, few of us have truly understood what the term "global village" means... Yet today, growing thousands of [people] in dozens of countries around the world are living the reality of the global village in personal, hands-on, interactive ways. Through the medium of networking and telecommunications technologies these students are for the first time learning to think of themselves as global citizens, seeing the world, and their place in the world, in ways much different that their parents" (Rogers, A. 1995)

As government and professional relationships in Asia and Pacific Rim expand during the next 5 years, Australian Universities must provide increasing technological opportunities for students studying at a distance. Distance Education today is the business of bridging language and cultural understanding through the use of technological tools. The use of interactive tools on the Web is one answer for overcoming distance, time and space in an ever shrinking world of academic resources.

Educational institutions need to prepare today for tomorrow's opportunities and part of that preparation is waking up to the fact that the way we communicate with each other is changing. The global village concept is becoming more and more a reality every day with a "global classroom" being virtually in the next room on a screen and half a world away. The Web is the interactive activity of the 90's and good AV materials incorporated for entertainment and teaching can bring it to life. If educational institutions and educators want to survive in the 90's and beyond, we need to get on the Web - AV - Internet train.

References

Book References:

(NBEET. 1996) National Board of Employment Education and Training (NBEET) (1996). Education and Technology Convergence. A survey of technological infrastructure in Education and the Professional development and Support of Educators and Trainers in Information and Communication Technologies. Commissioned Report No. 43. p. 94

Proceedings references:

(Watson, D. M. 1995) Watson, D. M., (1995). "The Impact of Live Audio Visuals on the WEB and Business" in R. Debreceny & A. Ellis (ed.) AusWeb 95 - Innovation and Diversity The World Wide Web in Australia, Norsearch Ltd., Lismore, NSW, p. 271-275

(Watson, D. M. 1996) Watson, D. M. (1996). Making IT Happen in Real Time: A review of audio technologies for teaching health management topics via the Internet. Health Informatics Conference Proceedings, HIC 96, Melbourne, VIC. Australia (in press)

Journal references:

(Perkins, D.N. 1991) Perkins, D.N. (1991). Technology meets constructivism: Do they make a marriage? Educational Technology, May, pp. 18-23.

(Watson, K.K. 1993) Watson, K.K. (1993). Computers and Collaborative Interactions. Australian Educational Computing, July, pp. 293-295.

World Wide Web references:

(Real Audio - http://www.realaudio.com) Real Audio Homepage, http://www.realaudio.com

(Rogers, A. 1995) Rogers, Al (n.d). Global Literacy in a Gutenberg Culture. arogers@bonita.cerf.fred.org

http://gsn.org/gsn/article.gutenberg.html

The Virtual University: Making It Happen

Matthias O. Will
Institut für Informatik
Universität Freiburg
e-mail: will@informatik.uni-freiburg.de

1 Introduction

Teaching at universities is mostly constrained to using blackboards, overhead transparencies and sometimes flip-charts as basic tools. In order to consolidate the course material, and as a preparation to written or oral exams, students have been using course notes, textbooks or copies of transparencies for decades.

However, with the advent of powerful hardware and network technology, the concepts of computer-based education, which make it possible to follow university curricula independent of time and space, is now becoming a reality. Conversely to most approaches to this topic, which either focus on developing authoring tools for the realization of CBT courses, or on remote broadcasting of lectures, we are working on merging these two concepts in order to make full use of new technologies as a supplement to traditional teaching at university level. As an example, we are working on developing an educational environment in the area of algorithms and data structures on the basis of Hyper-G, a powerful, second-generation hypermedia system.

In section 2, we will describe a few scenarios for the enhancement of traditional teaching methods wee were considering for the development of a prototype at our institute (see section 3). We conclude by discussing some crucial aspects for our future research (see section 4).

2 Enhancing Traditional University Education: A Starting Point

The traditional approach for university education can be seen from two viewpoints. Depending on the course type, the professor first makes an outline to find out what kind of material he will use, followed by gathering all kinds of scientific material (books, journal articles) on the relevant topic to prepare each course unit. When teaching the course, he uses any combination of his own notes, overhead transparencies or slides. Of course, if the course has been taught before, only little effort is necessary to keep the course material up to date.

On the other hand, the student is often given an outline of the course along with a list of relevant course material, which can either be bought or borrowed from a library. Sometimes, copies of the transparencies or lecture notes are handed out. Thus, students can benefit from a potentially large number of written material to later review the course or lecture at home. Additionally, tutorials are often offered to support the student in his learning process.

This technique has worked well for decades and therefore can be used as a good starting point in order to keep the advantages for both parties while guaranteeing a smooth transition to tomorrow's university education.

Since the basis for university education are lectures and printed materials, a first scenario would be the automatic conversion of lectures into multimedia documents ([OBa95], [BaO96]), which can then be inserted into a Web server and combined with digital library access, thus making the lecture available publicly unconstrained by time and space. This would solve the problem of lectures having to travel frequently to deliver the same course at different universities.

By adding multimedia documents (e. g. animations, simulations) to digitized texts, interlinking them and storing them into a multimedia database server, it is possible to benefit even more from computer-specific resources.

The user of such a system is mainly absorbing information. but tutorial support is missing, e. g. by answering the student's questions or having him solve problems. The latter is possible by integrating multiple-choice tests, however, these are not suitable for more complicated problems and do not support the student in his strength and weaknesses. This will be discussed in more detail in the next section.

3 A Prototype for Tomorrow's University Education

The development of computer-based courses ist most often done using special authoring software. Usually, an interdisciplinary team of content providers, media experts and designers join their efforts to produce a multimedia package, which is then sold to the general public as a CD-ROM. The time and costs for producing such a package are substantial: everything has to be developed from scratch since no parts or modules can be used from previously produced packages. Also, it is not possible to modify, enhance or adapt such a course to one's needs.

However, with the amount of educational resources freely available on the Internet, we favour an open, modular approach, which can be compared to the object-oriented software-engineering approach, where previously implemented software modules can be easily reused. Furthermore, in order to facilitate the development of course material, it is desirable to be able to focus on content instead of the graphical presentation of a specific topic[+].

As a way out of the dilemma, our approach is to combine the documents of an electronic library with on-line lecturing, to add multimedia components and tutorial elements, which are indexed upon insertion to enable full retrieval, and to integrate everything into a powerful, networked hypermedia system such as Hyper-G ([Ma96], [DH95]). Thus, if provided with graphical interfaces for the course developer and the student, we are able to produce modular, interactive educational environments at little costs where feedback between the system and the user ensures that the benefit for the student is substantial if used in conjunction with traditional studying techniques.

As the basis for our prototype, we took a commonly used textbook in undergraduate curricula, whose postscript version was partitioned in sufficiently small sections and then inserted into the server. A number of freely available computer simulations were enhanced, provided with a user-friendly interface and linked to the relevant book section. If suitable, we provided links to the appropriate parts of previously stored on-line lectures. The assignments were also added along with the solutions so they could be retrieved later.

To register, students fill out a form from their Web browsers to be able to access the educational environment after identification. The list of students is automatically kept up-to-date and can be used later for management purposes. Identified students can add own data or references or make annotations to any kind of document, thus creating a personalized view of the environment.

In order to improve navigation between various parts of the system, it is crucial to introduce typed links to inform the user about the document type and additional information about the document being accessed when following a link.

Since undergraduate computer science courses focus on teaching basic algorithms and data structures, it is important to include a programming environment, which enables the students to train their programming skills and to visualize and analyze the output of their efforts with the help of a visual debugger. Thus, if an assignment is completed, the code could be automatically extracted and linked to the solved assignments. Since the correction of theory assignments cannot be automated, the solution, which is written with the help of a text editor, is sent to the system upon completion. The solution can automatically be forwarded to the assigned tutor, or either a link is generated on the fly between the tutor's collection associated with a spefic student and a completed assignment. Comments from the tutor can be appended via the annotation mechanism.

Concerning the tutorial aspects of a system, i. e. the ability to ask and answer questions, the feedback for questions having been asked before can be automated by accessing a database which contains the corresponding answers. All other questions could be forwarded to the tutors via e-mail if they are not currently logged on, or via a text window which the tutor can access via the user interface. If an answer can be provided on-line, an audio and/or video connection can be established between the tutor and the student. The tutor also types in the answer in another window so the question and answer pair can automatically be inserted into the database.

In order to enhance interaction between course participants (including tutors and lecturers), a record can be kept up-to-date where personal data is stored along with a record of past interactions so that it is possible to consult others who are working on a similar problem or who are possibly able to answer a question on a specific topic. However, information about personal achievements should only be made available upon agreement of the contacted participant. Contact between users could be established via e-mail or other communication facilities. An advanced setting would be the integration of CSCW tools to enable users to jointly work on a specific task.

4 Conclusion

In order to build an educational environment which is truly beneficial to the participants, extending traditional techniques in education by the use of computer technology, it is important to gain information about the acceptance of such a system. Our first experience clearly shows that it depends on its user-friendliness, the interaction facilities and its technical performance. An interdisciplinary approach involving cognitive and educational scientists is certainly beneficial to find out about the possible enhancements, along with feedback and suggestions from course participants. Although we currently focus on the field of algorithms and data structures, we believe that hypermedia educational systems can be beneficial for a large number of sciences. Thus, our aim is to develop an open, extendable system which can easily be adapted for the use in other disciplines, tailored to the needs of both lecturers and students.

References

AOS91
F. Augenstein, Th. Ottmann, J. Schöning: *Ein typ- und regelgesteuertes Autorensystem*. In: H. Maurer (ed.): *Hypertext/Hypermedia '91*, pp. 25-33, Informatik-Fachbericht Nr. 276, Graz, Austria, 1991.

DH95
Wolfgang Dalitz, Gernot Heyer: *Hyper-G. Das Internet-Informationssystem der 2. Generation.* dpunkt-Verlag, Heidelberg, 1995

Ma96
H. Maurer: *Hyper-G now Hyperwave. The next generation web solution.* Addisson-Wesley Publishing Company, Harlow, England, 1996

OBa95
Th. Ottmann, Chr. Bacher: *Authoring on the Fly*. Technical Report, Universität Freiburg, October 1995

BaO96
Chr. Bacher, Th. Ottmann: *Tools and Services for Authoring on the Fly*, EDMEDIA '96, Boston, June 1996

Ott95
Th. Ottmann: *Hochschullehrer als Autoren und Nutzer von Lernsoftware*, Universität Freiburg, December 1995.

LMa94
J. Lennon, H. Maurer: *Lecturing technology, a future with hypermedia*. Educational Technology, pp. 5 - 14, April 1994

StHay91
J. T. Stasko, D. Hayes: *XTANGO Algorithm Animation Designer's Package*, College of Computing, Geogia Institute of Technology, Atlanta (1991)

Timely Tricks for Web Language Teaching

P. Fawn Whittaker
Language, Literature & Communications
Brigham Young University Hawaii Campus
United States
whittakf@byuh.edu

How can you easily produce inexpensive interactive language learning materials? With expensive, complicated software and limited budgets, most busy language teachers turn back to the traditional chalkboard and perhaps an Overhead Projector. However, with low-cost educational internet and server access provided by the school, a camcorder or other digital camera, basic html knowledge with free software, and some sound and video freeware or shareware downloaded from the internet, the language instructor can produce endless interactive language learning exercises for not only his/her local students but for students around the world.

Hardware Needs

The computing personnel of your school should provide part of the basic hardware needs: a computer with server capabilities. This could be a PC, UNIX, or Mac with WEBstar. You should have remote access to a folder on that server, at least 500 MB of space there, and a password into that folder on the server. And with a browser such as Netscape Navigator, you can easily access the world of language teaching already available on the WEB, to which you will soon be contributing via your server folder.

You will also need a computer from which to access the world's WWW language pages and produce your own pages. This computer should have A/V capabilities so that you can listen to and produce sound as well as pictures. You should also have a microphone to connect to your A/V computer for sound production.

Another piece of hardware that will enhance your work is a digitized camera (averaging $650 at present) or a camcorder. If you choose to use your personal, or the school's, camcorder connect the cables (usually provided at initial purchase) from the RCA, and possibly S-Video, ports of your camcorder into the same of your A/V computer. Insert a blank video tape (any version is fine: VHS, beta, Hi-8, etc) to record.

Software Needs

Software requirements are also basic. For html (hypertext markup language) production, the basically free software Simpletext (in any version) will suffice. Or you can download a myriad of other html editors for free by accessing their titles via Net Search. The search engine Yahoo at <http://www.yahoo.com/> offers some good sources under "WWW" and "Authoring".

For listening to and producing sound, you will need a sound editing program. When you try to listen to a sound, but the program is missing on your hard drive, you will be informed what program to obtain. A good quality freeware sound editor that you can download off the internet is "SoundEffects". This, together with several other commonly used sound programs, is found at the following location: <http://www.wavenet.com/~axgrindr/quimby4.html>.

As for capturing pictures from your camcorder, you will need Apple's Video Monitor (requires only 77K on your hard drive) or Avid Videoshop (which is commonly packaged with A/V PowerMacs and requires 1.5 MB on your hard drive). If you use a digitized camera, you will need the software which comes with the camera. You will also need software to convert your PICT image into a GIF or JPEG image, which you can do with shareware entitled "GraphicConverter" (which requires 629K) or with the more expensive Adobe Photoshop, available at educational prices (which requires 8 MB). With this software you can capture and convert images to use as illustrations in your WWW language lessons.

Finally, if you are preparing lessons in a language that requires fonts that are vastly different from the norm, you may need to purchase language software, such as the Chinese or Japanese or other language kits through Apple. If it is simple diacritics that you need for Western European, Native American, or Pacific Island languages, these are expressed using metacharacters in your html which are listed in most html texts.

Production

Now that you have the hardware and software that you need, you can begin production. Begin by outlining the language lessons which you would like to incorporate: pronunciation, greetings, numbers, pronouns, verb tenses, modifiers, etc. Make a graphic outline of what you want to make available.

When you have a basic outline, you are ready to begin typing in your lessons. For a pronunciation presentation, for example, you might want to include the alphabet and words, phrases, and sentences with the pronunciation point in question. Upon typing in the content, your next step is to record the sounds you have typed which you want the students to listen to and repeat. Record small chunks of sound in "mono" rather than stereo, at 11Khz, and at 8 bits so that the downloading time is kept at a minimum. Save each of these sound files with a brief name that you can remember, ending with ".aiff". An example is <teuana.aiff>. As you save the sound files, it is a good idea to save them all into a separate sound folder within your language folder with the name "sounds" or a comparable title.

If you wish to insert a picture to illustrate the language point, take a "picture" with your camcorder of the desired object or picutre. For example, if you want a picture indicating the position of the tongue for the /shr/ sound in Chinese, you will videotape the previously prepared diagram, then connect your video RCA or S-Video cabling from the camcorder into the computer, rewind to the desired position on the video tape, and "capture" the picture into the computer with the video software that you have selected. This picture will be saved as a PICT file, but must be converted to a JPEG or preferrably GIF file--GIF because the format is more widely accepted. A low to medium quality conversion should be acceptable on the internet and will keep the size of your image file small. After you have converted your image, give it a short title, ending with ".GIF" or ".JPEG" according to your selection. An example is <Haus.gif>.

Now that you have the text, the sound, and the pictures, your job is to link them together in your html document so that students can view the text with an illustrative picture or diagram, press on a word, listen to it, repeat it, and even record their own perfected versions on their own diskettes. You will have produced a masterpiece of language interactivity that your students can enjoy and work with long after class has been dismissed. You will have created their personal home tutor, which students anywhere can enjoy.

Samples of the presenter's WWW language learning lessons in various Less Commonly Taught Languages can be found at the following location: <http://lc.byuh.edu/Samoan/Samoan.html> and <http://lc.byuh.edu/Kiribati/kiricoco.html>. These also apply netforms, which software can be obtained for under $100 from Maxum Development Corporation, and image mapping, which can be obtained as freeware from the internet by searching under "Webmapping". Illustrations of how this software can be used together with the basics described in this paper to increase WWW language teaching effectiveness, using current daily news and up-to-date technological tricks will be demonstrated to attendees at the conclusion of the presentation.

The Bible and Talmudic Literature on the Web

Itzhak Yarhy
Snunit Edu. Info. System
Hebrew University of Jerusalem
Israel
itzyar@www1.huji.ac.il

Ofer Margoninski
Snunit Edu. Info. System
Hebrew University
Israel
omargo@www1.huji.ac.il

David Rashty
Snunit Edu. Info System
Hebrew University
Israel
rashty@www1.huji.ac.il

Prof. Nava Ben Tzvi
Science Teaching Center
Hebrew University
Israel
nava@mail.snunit.k12.il

The Jewish tradition is based on two laws: The Written Law and the Oral Law. According to Jewish belief, both laws were given together at Mount Sinai. The Hebrew Bible, compiled during the first millennium, BC , and constitutes the written law. The Talmudic Literature, composed of the Mishna, Tosefta and the Talmud (the Palestinian Talmud and the Babylonian Talmud) is the first codification of the Jewish "Oral - Law". It was compiled during the second, third and fourth centuries, AD .

The ability to search the Bible and the Talmudic Literature in 'full text' (i.e. - search for a specific word or words that appear anywhere in the text) is, of course, of great importance to students and scholars. Until recently, the only way to conduct such a search was to use a Concordance - a voluminous book in which all the words appearing in the Bible are listed in alphabetical order. For each word the places in the Bible where it appears are listed. In recent years several CD's that enable full text search of the Bible became available on the market [Bar Ilan Univ. The Responsa], but there were still no widely available, simple to use tools, enabling a full text search of the Bible on the Internet.

The Talmudic Literature quotes the Old Testament frequently, in order to prove or disprove claims or to bring examples. Most versions of the Talmudic Literature designate the chapter (but not the verse) from which each quote is taken. The Bible, compiled a long time before the Talmudic Literature, of course doesn't have cross references to those quotes. Thus another desired feature in databases of the Bible and the Talmudic Literature, is a thorough cross referencing between the Bible and the Talmudic Literature, based on the quotes described above, and the ability to easily follow those cross references.

We created a database containing the Bible and the Talmudic Literature, in Hebrew (ISO 8859-8) [Nussbacher & bourvine, 1993][Snunit, 1995] on the Web, that has the following features:

1. Full text search of the Bible and Talmudic Literature
2. Easy to use, Hebrew interface for a graphical browser working on a variety of operating systems.
3. Cross references from the Talmudic Literature to the Bible and vice versa, implemented through

HTML links.

The user accessing the database (at http://www1.snunit.k12.il/kodesh/kodesh.html) can descend through a series of menus, reflecting the hierarchical structure of the Bible and the Talmudic Literature, and read a specific chapter. He can also search the Bible or the Talmudic Literature for specific keywords, phrases, or keywords connected with Boolean operators, by using an HTML form interface. In response to his query, the user receives an HTML page [Figure 1] in which the list of relevant chapters that answered his query appears. The list is ordered according to the traditional ordering of the books in the Bible. For each chapter, all the verses containing the words searched by the user are represented as HTML links. When pressing one of those HTML links, the user is presented with the chapter itself. The relevant verse appears on top of the browser's text window.

In the database, each chapter is represented as an HTML page. The Talmudic Literature texts quote often from verses in the Bible. In the database those quotes are represented as HTML links pointing to the exact verse from which the quote was taken. On the other hand, when a specific verse is quoted in the Talmudic Literature, it appears highlighted as an HTML link in the Bible. Pressing this link brings up to the user a different version of the chapter[Figure 2], which includes HTML links to the places in the Talmudic Literature that quote this verse. Thus the user can easily view the commentary from the Talmudic Literature available for each verse.

Architecture and Implementation:

The database itself is simply a collection of HTML pages, one for each chapter. The files were indexed using FreeWAIS 0.5. A 'WAISGate' [Tim BL & H. Frystyk 1995], is responsible for the interface between the WAIS server and the client's browser. Special modules were added to the WAISGate in order to support Hebrew search and return the list of matching verses according to the conventional order of the books in the Bible. We are currently replacing WAIS with 'Inter-Text', a commercial indexer, especially adapted for Hebrew, made by SPL.

As mentioned above, the Talmudic Literature contains a lot of quotes from verses in the Bible. For each quote, the chapter from which the quote was taken is mentioned, but not the verse. Also, many times the quotes are not exact: some words are spelled differently or omitted all together and it isn't clear where the quote ends and the Talmudic text resumes. A special algorithm, implemented in Perl, was developped in order to over come those problems and find the exact verse quoted. The algorithm rated all verses in the chapter according to how much they matched a given quote, and then chose the verse with the highest score as the one from which the quote was taken. If it seemed that several verses matched a given quote, the length of the quote was doubled, and the score for each verse recomputed. After the quoted verse was found, a hyper link pointing to it was made in the Talmudic text. Another hyper link, pointing to the Talmudic text, was made near the relevant verse in the Bible.

Conclusion

We have implemented a database of the Bible and the Talmudic Literature on the World Wide Web. The database enables the user to easily reach the chapter he is interested in, and view it in a way that is pleasant to the eye. By using HTML links embedded within the text, the user can easily move between a specific chapter and the commentaries relating to it in the Talmudic Literature. The user can also search the database for any words or phrases appearing in it.

We believe that databases, widely available to the public, should be one of the main constituents of the 'useful' portion of the World Wide Web. Here we have demonstrated how a rather advanced database can be built by using and modifying widely available freeware and combining it with some original ideas and algorithms. We also demonstrated how an 'electronic version' of a book can be produced on the Web, and how such a version can support advanced browsing features which would never be available in a conventional book.

References

[Bar Ilan Univ, The Respnsa] Bar Ilan University, The Responsa Project, URL: http://www.biu.ac.il/BIU/Jewish/responsa.html

[Nussbacher & Bourvine 1993] H. Nussbacher, Y. Bourvine, RFC 1555, "Hebrew Character Encoding for Internet Messages", URL: http://www.internic.net/rfc/rfc1555.txt 12/23/1993

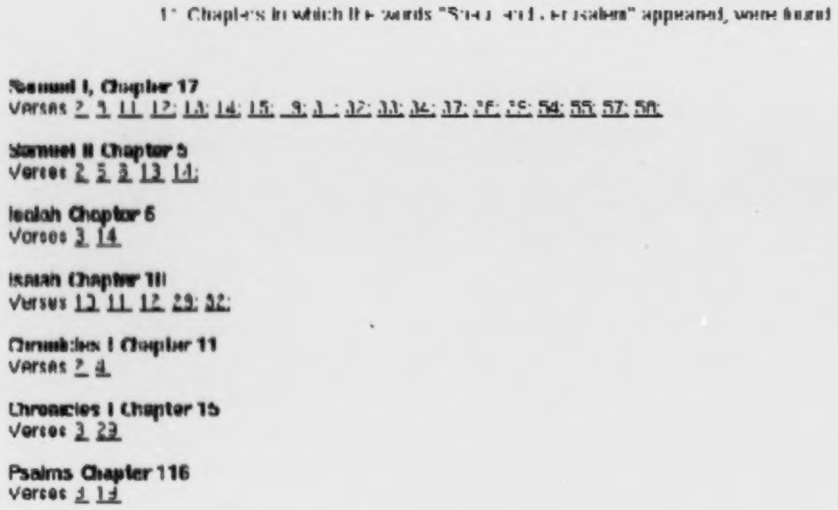
[Snunit 1995] How to read Hebrew on the World Wide Web, URL: http://www1.snunit.k12.il/snunit/heb.html

[Tim BL & H. Frystyk 1995] Tim BL and Henrik Frystyk, The WAIS - WWW gateway URL: http://www.w3.org/pub/WWW/Gateways/WAISGate.html

Leslie D. Cuff, Commercial Hypertext Publishing - "Electronic Books Using Trails and the Author-Publisher-Reader Model", World Wide Web Journal, Fourth International World Wide Web Conference Proceedings,December 1995, Published by O'Reilly & Associates, Inc.

[Figure 1] URL: http://www1.snunit.k12.il/shortfig1.html

[Figure 2] URL: http://www1.snunit.k12.il/shortfig2.html



[illegible] Chapters in which the words "Shaul and Jerusalem" appeared, were found

Samuel I, Chapter 17
Verses 2 8 11 12 13 14 15 19 31 32 33 34 37 38 39 54 55 57 58

Samuel II Chapter 5
Verses 2 5 6 13 14

Isaiah Chapter 5
Verses 3 14

Isaiah Chapter 10
Verses 10 11 12 29 32

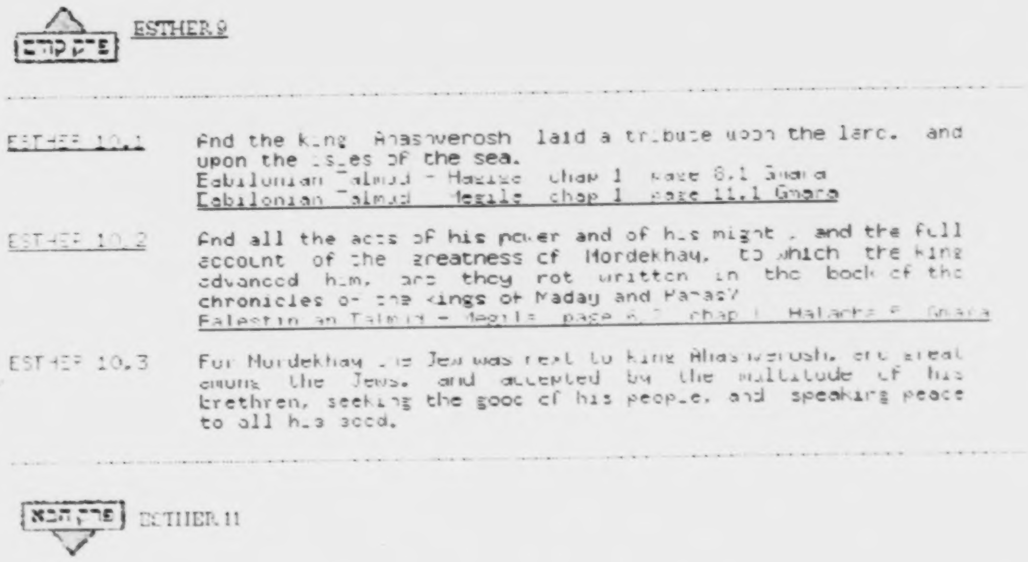
Chronicles I Chapter 11
Verses 2 4

Chronicles I Chapter 15
Verses 3 29

Psalms Chapter 116
Verses 3 19

Figure 1: The HTML page that the user receives in response to his query "Shaul and Jerusalem". The chapters that answered the user query are listed according to the traditional ordering of the books in the Bible. For each chapter, all the verses containing the words searched by the user are represented as HTML links.

BEST COPY AVAILABLE



ESTHER 9

ESTHER 10.1 And the king Ahashverosh laid a tribute upon the land, and upon the isles of the sea.
Babilonian Talmud - Hagiga chap 1 page 8.1 Gmara
Babilonian Talmud Megila chap 1 page 11.1 Gmara

ESTHER 10.2 And all the acts of his power and of his might, and the full account of the greatness of Mordekhay, to which the king advanced him, are they not written in the book of the chronicles of the kings of Maday and Paras?
Palestinian Talmud - Megila page 6.2 chap 1 Halacha 5 Gmara

ESTHER 10.3 For Mordekhay the Jew was next to king Ahashverosh, and great among the Jews, and accepted by the multitude of his brethren, seeking the good of his people, and speaking peace to all his seed.

ESTHER 11

Figure 2: The HTML page of Esther, chapter 10, in the bible database. The underlined lines are the hyperlinks to relevant commentaries in the Babilonian and Palestinian Talmud. (The original HTML page, is ofcourse, written in Hebrew, this is a translation to English of the page.)

BEST COPY AVAILABLE

The Platform of Corporate Development and Management CAI Software on The WEB

Jia Zhuosheng, Wei Huiqing
100044
Compute Center
Northern Jiaotong University
Beijing. P.R. China
email: jia@center.njtu.edu.cn, hqwei@center.njtu.edu.cn

Now a day, the development, application and management of the Computer Aided Instruction (CAI) course software is comparatively disperse in our country. A lot of university and institute develop the same application course software bye themselves, seldom exchange and less corporate. They do a lot of duplicate job in lower layer, and waste much time of software developer. The course software platform of development and application is deferent, it is not convenient for users. Even though they develop some course software cooperatively, they can not exchange each other in time due to transport and communication etc. And the more department take part in, the more difficult management and corporation.

In view of the case, I designed and built a open environment for development, application and management CAI course software in Web server. The people in different institute and department can contact each other by network. It Concentrates developers on a network development platform. On this platform, they may take part in deferent develop group. The developer can register their preferring develop group after they are certified, they can participate in developing course software about this group together. Each developer can not only watch progress about this group in time, but also can apply to group manager for a part of development of course software.

By this way, the group manager will dispatch all matter. Avoiding duplication development, raising working efficiency. the managers always inquiry the newest trends and collect metrical form internet about his subject Under the management environment, and supply it to information database of development environment. So the developers needn't spend much time to inquiry message form internet, only entry the development environment, they can see updated database conveniently. In this way it can make many people develop a high level general CAI course software in short time. Because the run platform and user interface are same in the Web server. So it is conveniently to transplant and application.

This system link the user interface with develop interface by management module, the found problem by user can reflect to the develop environment immediately. Developers can know the situation of utilization in time, in order to perfect and update the course software .

Beside offer its own aided instruction function, the user module can let students organizing a different study group, discussing problem, asking teacher, posting letter and call system manager directly advancing exist problem and improve method. The system management module responsible organize and coordinate CAI course software developer, collect information, manage user login, gather user's opinion and statistic, account etc.

So we fully utilize web service to solve the shortage of CAI course software development, application and management, make it always follow the developing of the world advanced level, and let CAI education break the wall of collage, offer service for all the country and society.

The Internet & The WWW: Friend or Foe?

Dr M J McAlister
Senior Lecturer in Computing
University of Sunderland

P Comer
Senior Lecturer in Computing
University of Sunderland

Miss K Hadley
Lecturer in Law
University of Sunderland

This panel discussion will consider the Internet and its future with reference to relevant legal issues and risk analysis.

The Internet grew out of a contract to the Advanced Research Projects Agency (APAR) of the US Department of Defense. A network called APARNET was developed at Bolt, Beranek and Newman which connected university, military and defense contractors which would aid the communication process between researchers at the various institutes and provide a means of estimating how communications could be maintained in the event of a nuclear attack. The network was expanded to allow remote login, file transfer capabilities, electronic mail and mailing lists to be maintained. At the same time other networks and protocols for transmission of information were being developed and a method communication became a central issue. The networks were connected via gateways, routers and hubs which passed the traffic from one network to another.

The Internet provides a range of services, some of the best known are, InterNIC, Finger, Internet Relay Chat, Mud Netfind, Talk, Whois, X-Window system, remote login, File Transfer Protocol (FTP), electronic mail, Wide Area Information Server (WAIS) and the World Wide Web (WWW). The WWW is an Internet service which provides an advanced browsing system that allows many computers to access a hypermedia environment. Originally designed at CERN to allow particle physicists to share their knowledge with on another around the world, the WWW has expended to include a number of organizations both academic and commercial who are taking this opportunity to publise themselves and their products world wide. Examples of its use include the setting up of news groups, advertising of conferences, reproduction of newspapers, electronic journals, distance learning courseware, freeware/shareware which can be downloaded, bulletin boards, advertising of new books which can be bought by returning a pin number and advertisement of academic/commercial institutions. With such an expanse of information being passed around a control mechanism must be considered. At present the World Wide Web Consortium (W3C) a volunteer organization is responsible for developing common standards for the continuing evolution of the WWW.

From the perspective of users and designers the WWW has been seen to become unmanageable, insecure and in some respects being used for illegal purposes. From this basis the panellists will be considering the following issues with relevance to the UK and subsequently to other countries throughout the world:

1. What security devices can be implemented to control what is published on the WWW?
2. Most people will be well aware of the risks from viruses but what about other potential risks?
3. How many organizations have the resources to perform an in-depth risk analysis?
4. What safeguards are in place to frustrate hackers?
5. What are the implication on intellectual property with regard to data protection, copyrights, patents and hacking?
6. What are the implication for criminal offences including theft and obscene publications?

7. What are the implications on contracts with regard to negligence, nuisance and deformation?
8. What are the implications on consumer law?

Friedrich-
Alexander-
Universität
Erlangen-
Nürnberg

Dermatologische Universitätsklinik

(Chairman: Prof. Dr. G. Schuler)

Hartmannstr. 14, 91052 Erlangen, GERMANY

Email: bittorf@derma.med.uni-erlangen.de

Telephone: +49 - 9131 - 85 - 27 27, -68 37

Fax: +49 - 9131 - 85 - 38 50, -27 28

Dipl. Inf. A. Bittorf

WebNet 96

World Conference of the Web Society

October 15-19, 1996, San Francisco, USA

How to set up and maintain an image database on the WWW efficiently

Andreas Bittorf

University of Erlangen, Germany

Abstract:

In various medical specialties as well as in other disciplines like marketing, chemistry, biology, etc. there is a great demand for reference images organized in databases. Via the WWW those images can be made available to users everywhere.

After the tutorial the participants will be able to decide how to handle their own images because many examples will be shown, demonstrating the effect of manipulation like compression, gamma correction, as well as the change of resolution and color depth. Further, image acquisition, quality, formats, and color aspects will be addressed. Storage and retrieval mechanisms will be discussed and the effects of a proper design of the database and description system will be shown. For the WWW-interface alternatives will be given, considering access mechanisms provided, integration of distributed data,

server as well as network load, and maintenance. Each of these issues will be discussed in detail using examples. The last part of the tutorial will be a case study, the Dermatology Online Atlas (DOIA) project in Erlangen.

After the tutorial the participants should be able to decide how to set up a WWW based image "database", using notes and bookmarks.

Table of Contents

1 Introduction 12 Images 12.1 Basics 12.1.1 Image Types 12.1.2 Image Acquisition 22.2 Manipulation 42.3 Colors 52.4 Compression 62.4.1 Techniques 72.5 File Formats 72.6 Plug-Ins and viewers 93 Database 93.1 Database Types 103.1.1 Relational DBMS 103.1.2 Object Oriented DBMS 103.1.3 Text Oriented DBMS 113.2 Image Description 113.2.1 Key Systems 113.2.2 Conceptual Graphs 124 WWW Interface 124.1 File System versus Database 134.2 General Architecture 144.2.1 Forms interface 144.2.2 The Common Gateway Interface 164.2.3 Database server demon 164.3 Tools 164.3.1 GSQL 174.3.2 Internet Information Server 204.4 Database versus Information Retrieval 224.5 Interface Design 224.6 Integration of distributed data 224.6.1 Access times 234.6.2 Consistent view 235 Case study: DOIA 245.1 Images 245.2 The local database 255.3 Mapping the local Database to the Web 255.4 Acceptance and user profile 256 Summary 267 Acknowledgments 268 References 27

1. Introduction

The increase in computing power during the last few years has enabled personal computers to deal with graphics very efficiently. To date graphics cards are available for a few hundred dollars that can display images with a resolution of over one million picture elements (pixels) and 16,8 million colors. This makes image processing or at least display available to the average end user. Due to this during the last years more and more image-applications for personal computers evolved. They range from online shopping catalogs to spatial information systems, from art galleries like „the Louvre" to medical research databases, e.g. the „visible human project" or the „dermatology online atlas". Images are invaluable sources of information: „An image is worth a thousand words". All issues related to images will be discussed in the first section.

However it is quite difficult not to get lost among thousands of images. Therefore, there must be a structured way to access images. That is where databases may help. However, there are many different database systems, paradigms, features, etc. What kind of database is needed? Would not an easy text search be everything one would like to have? Think of information retrieval functionality. What about image manipulation features? Wouldn't it be great if the database could provide all this? How could images be described so that not only you but also others may use the database? What kind of description is most suitable?

Can the database be connected to the Web to be useful to others? What tools are available? Are there „out of the box" solutions provided by vendors? How can the local database be interconnected with other databases to create a global information infrastructure related to your research field? Are there advantages of a WWW front-end even when not providing public access to the database?

Finally, we will report how we answered all these questions when setting up our Dermatology Online Atlas (DOIA) and discuss problems we came across and solutions we have found in each of the above mentioned categories.

1. Images

This section is divided into six major parts. In the beginning we will cover some basics, continue with manipulation, compression techniques, aspects of color in digital images, discuss file formats and finally have a short look at Plug-Ins for Web-browsers.

1. Basics

2. Image Types

When talking of images the following categories can be distinguished:

- Object description
- Pixel graphics
- Mixed types

An object description describes not the image as a whole but the objects contained. Well known file types of this category are the Computer Graphics Metafile (CGM), Word Perfect Graphics (WPG) or Windows Metafile (WMF). For computer aided design there are extensive specifications like the Initial Graphics Exchange Specification (IGES) or the successor Standard for the Exchange of Product model data (STEP). Description languages like postscript fall also in this category. Images stored in these formats are very useful as many operations can be performed like scaling, and moving or deleting an object.

In many cases, however, an image can hardly be described with objects. Photographic images for example are usually described with pixel graphics. The word pixel comes from PICture ELement, sometimes pel is used instead. A pixel graphic could be viewed as if a grid was placed over the image. The cells of the grid are the pixels, the finer the grid the better is the representation of the image. Color values of the cells are calculated as the weighted averages of the colors present in the cell.

Note: A pixel graphic is inherently „lossy" due to the analog/digital conversion. Consider this when thinking of lossless and lossy compression.

Each cell is represented by a value that characterizes the color. Pixel images tend to allocate a lot of memory and need much computing power. For example an A4 page (21x29,7 cm) scanned at 158 dots per centimeter (400 dots per inch (dpi)) with true color (3x8bit) occupies about 44 MB of memory. High end scanners and digital cameras may produce images with a size of several hundreds of MB.

Mixed types, that combine pixel graphics with objects, are very useful, for example think of textual comments or marking regions of interest. Unfortunately only few image types support the storage of object and pixel graphics in one file.

Pixel graphics may be converted to object graphics and vice versa of course with strong limitations. In this section we will concentrate on pixel graphics, however the statements of the other sections hold for object graphics as well.

1. Image Acquisition

There are many ways to get digital images. We have selected some of the most important:

- Kodak Photo Compact Disc (PCD)
- Scanner
- Video + Framegrabber
- Digital camera
- Imaging devices

The Kodak Photo Compact Disc (PCD) [1] is a "write once, read many times" (WORM) medium. It is cheap in production and durable. For a Photo CD Master Disc 24x36 mm films are scanned with highly calibrated devices at 2170 dpi and 3x12 bit color resolution (RGB). The images are stored in the YCC color space, which is compatible to CIE-Standards; this implies that colors can be reproduced correctly.

Available are different formats:
- Professional CD Master Disc: 25 images, 6 different resolutions
- Photo CD Master Disc: 100 images, 5 different resolutions
- Photo CD Portfolio Disc: 700 images or 60 minutes of sound or a combination
- Photo CD Catalog Disc: Up to 4500 images in low resolution
- A Photo CD Diagnostic Disc with up to 8192 * 12.288 Pixels was announced, however, the plans have been dropped

The resolutions available on PCDs are shown in Table 1.

Name	Resolution	Memory (MB)	Comment
Base/16	128 * 192	0,070	Index Print
Base/4	256 * 384	0,28	Preview
Base	512 * 768	1,125	For CD-I
4 Base	1024 * 1536	4,5	Documentation and HDTV-Standard
16 Base	2048 * 3072	18	Digital prints, ¼ A4
64 Base	4096 * 6144	72	Digital prints > A4, exposure
(256 Base	8192 * 12.288	288	Medical diagnostics)

Table 1: Resolutions offered by the PCD

A slide or flatbed scanner may be a good alternative to the PCD, if images are needed immediately. However, we were not able to achieve the quality set by the Photo CD. We had especially problems with the dark parts of the image.

Quality was even worse with the Targa+ framegrabber card and a studio camera, that is resolution was limited, colors were not realistic, and lighting was difficult. This configuration was not acceptable for Dermatology. However, there are specific applications that use a video camera together with an optical zoom and a special cold light source with good results.

As digital camera we tested the Kodak DCS 420, which generates good images. We had some problems with the color adjustment. High quality digital cameras like the Kodak DCS 460 are still very expensive, whereas low quality cameras are quite cheap. Low quality cameras offer resolutions usually between fifty thousand pixels and half a million pixels using true color, that is somehow comparable to a video camera plus framegrabber card. High end cameras allow to use the usual equipment of a photo studio.

You may want to check the following aspects, when buying a (low end) digital camera:

- Autofocus?
- Zoom?
- Flash?
- Macro (what is the closest distance)?
- Image preview (without computer)?
- Resolution (maximum, different modes)?

- Image compression (quality, different modes, lossless)?
- Power supply, accu?
- Interfaces to computer (SCSI, serial, parallel)?
- Memory?
- Sensitiveness?
- Minimum exposing time (for moving images, for example sports)?

Low end digital cameras do not at all deliver image quality comparable to scanning photographic slides or prints. So the main purpose for using them may be the need to get images fast at moderate quality, for example when producing id-cards for public transportation with photographs.

Imaging devices usually use their own file formats. Sometimes they allow to export the image in some common format like TIFF. In medicine many different imaging devices are used [2], for example ultrasound, and computer tomography. Very important for these is the standard ACR/NEMA 2 and its successor DICOM ("Digital Imaging and Communications in Medicine") [3], which explains file formats and communication mechanisms. However, to explain DICOM is beyond the scope of this tutorial.

1. Manipulation

We call every change of the source data a *manipulation*. There are different purposes to manipulate an image, for example enhancement, compression, cropping, or change of resolution. Compression will be discussed in a separate section due to its essential importance in networking.

Most commonly used techniques for image enhancement are color correction, adjustment of brightness, contrast and gamma, labeling, and the usage of filters like sharpening. They can be performed from every state of the art image processing program, for example Photoshop. There are also very powerful free- and shareware programs for this like Paintshop Pro or Lview Pro.

Note: Manipulation should always be done before down scaling an image. Lossy compression should be applied only after all changes have been made.

Lossy compression should be applied after all changes have been made, because repeated decompression-plus-lossy-compression cycles will decrease the image quality. However there is one exception to this rule [4]. You can compress an image with JPEG, decompress it, make local changes and compress it again with exactly the same parameters as before. This will only affect the area close to the change.

To decide, in which resolution in space and color an image should be stored, have a look at the intended usage: Will it be display on a computer screen with about 70 dpi or is it print at 1200 dpi, is zooming required? How many colors are needed? With many image manipulation programs you can count the number of colors used in a specific image. Would you like to use greyscale images or color images?

Greyscale images usually use 1, 4, 8, 10, 12, and (16) bits per pixel (bpp). More than 8 bpp (corresponding to more than 256 levels of grey) are only used in specialized application areas like digital photography and radiology. For color images usually 4, 8, 15 (*direct color*), 16 (*high color*), 24 (*true color*), 36, and (48) bpp are used.

It is difficult to display a colored image realistically with 8 bpp or less. Due to this, *color palettes* have been invented which select the best suitable 256 colors out of 16,8 million for a given image. These color values will be stored for every image and the bit value for every pixel is just a reference to the color selection. Using an optimized color palette there will be hardly any difference visible between the 8 bpp image and the true color original. The quality of the image depends crucially on the selection of good palette entries, which is not always straight forward.

Another method, called *dithering*, is used to improve image display with a small amount of colors available. Dithering creates additional colors and shades from an existing palette by interspersing pixels of different colors, as colors cannot be mixed directly like in painting. The different colors can either be distributed randomly or regularly. The higher the resolution of the display, the smoother the dithered color will appear to the eye.

1. Colors

If we had a *controlled color system* in which a color could be specified and would be displayed on all devices in the same way we would not have to care about color. Unfortunately this is not the case. Color will be different everywhere: on the screen, in print, and when scanning. Monitors, printers, scanners, exposers, etc. have different characteristics. But let us start at the beginning.

Colored images on the computer screen are generated using a red, green, and blue (RGB) component. Due to this it is straight forward to store colors in this so called *RGB system*. If x is the number of pixels used per pixel, usually x/3 bits per pixel are used for each of the components. If x/3 is not an integer, some colors are assigned more bits than the others. For input the RGB system is also very convenient as charged coupled device (CCD) chips are commonly used with RGB filters. CCD chips are present in video cameras, digital cameras, and scanners.

Unfortunately this is different for printing applications because these use a subtractive color representation, whereas RGB is additive. Additive means that white is represented by the tuple containing the highest available values and black is represented by (0, 0, 0). In printing it is exactly the other way round, there the tuple (0, 0, 0, (0)) for the Cyan, Magenta, Yellow, (Black) (CMY(K)) represents plain white.

There are other color systems that represent colors, for example, with Hue, Saturation, Brightness (HSB), or Hue, Saturation, Value (HSV) for human convenience, and many others. Very important are the systems of the Commission Internationale de l'Eclairage (CIE), for example CIE L*a*b and CIE x*y*z that are based on an analysis of the human visual system. They allow exact specifications of colors, whereas RGB or CMYK do not. Unfortunately the color systems cannot directly be mapped to each other. There are colors, e.g. in CMYK, that cannot be represented in RGB (and vice versa). This implies that an image in print may look different from the RGB image on the screen. However, even a CMYK image on the screen may look different to the same CMYK image in print as different devices have different capabilities to represent colors.

And it is even worse.

Note: Two monitors may display the same image using the same color system and it may look completely different.

This may be due to the gamma values and the different colors of white (white point) and red, green and blue (chromacities). There are other minor factors as well.

A non linear correction has to be applied when displaying images, because the brightness produced by Cathode Ray Tubes (CRTs) is not proportional to the voltage, but to the voltage$^{\text{gamma}}$. Different monitors have different *gamma* values. To get the gamma values of a device, dithered colors (RGB) can be contrasted with not dithered colors. The gamma value can be determined by comparing them. For example, if a not dithered color looks as the same color dithered, the monitor's gamma has been corrected to a gamma value of one. To do gamma adjustment for a value not equal to one a precorrected test image may be used.

To display images correctly the gamma values should be used for a correction of the image. Some systems, for example MACs, allow to store a gamma value as system parameter everything displayed will be shown gamma corrected. With PCs, calibration tools could be used to set the system gamma. If no gamma correction is applied, images may look washed out or too dark.

Even with gamma correction, there may still be a problem: RGB = (255, 0, 0) just says that the pixel will be pure red, but what this is may differ from screen to screen. Good monitors provide the possibility to adjust the color temperature so that some compensation is possible. The color temperature is given in Kelvin. Common color temperatures are between 5000 K and 9500 K. With a higher temperature white looks more blue, with a lower temperature more reddish. Modern computer monitors usually allow to store different modes that can be use for different applications. A complete solution however needs a calibration tool.

1. Compression

Methods to reduce storage space are essential for an image database on the Web, because images occupy large amounts of hard disc space and network bandwidth is limited. The basic idea in compression is to reduce redundancy present in images, and if lossy, to throw away the least important data.

Usually „lossless" and „lossy" compression algorithms are distinguished. *Lossless* compression algorithms allow to rebuild the original data file from a compressed one, whereas *lossy* algorithms only allow to compute an approximation of the original. For many applications like word processors and source code only lossless compression techniques may be used. However, in the area of sound, video, and imaging, lossy compression algorithms are very important, especially considering that a digital image is usually only an approximation of the reality due to the analog/digital conversion.

There are a few facts about compression everybody should know:

- There is no lossless algorithm that compresses everything.
- Compressing the „same" image in a higher resolution will usually lead to a higher compression factor as redundancy is increased.
- As a rule of thumb: Lossless compression factors for images 1:2.5
- As a rule of thumb: Lossy compression factors for images: 1:10, 1:30 and higher

It can be shown with a simple proof [5] that there cannot be any lossless algorithm that compresses all files larger than a given size N: Lets assume there was such a program. This implies that each file of length N has to be mapped to a different file with length smaller N. There are 2^N files of length N but only 2^N-1 of length smaller N. So there cannot be a bijection.

Increasing the resolution will create additional pixels that usually will have a similar color as the pixels nearby. Due to this only little information is added, the redundancy increased and a higher compression ratio achieved.

Of course there are special applications classes that compress much better than 1:2.5, for example constructed images like presentation slides, or postscript files, which may be compressed by 1:10 or better. However, the average photographic image will only compress lossless with a factor of 2.5 or less. With our dermatologic images we got factors of about 1.5 with Lempel-Ziv-Welch (see below) compression and about 2.2 with the compression implemented in PNG (see below).

Lossy compression throws away information. Due to this, much higher compression ratios can be achieved compared to lossless compression. Lossy compression techniques have been designed for the human observer. Usually there is no visible difference between original and compressed image at a compression factor of 1:10 for printed images and 1:30 for monitor images. Although there is no difference for a human observer lossy compressed images usually cannot be used for pattern analysis purposes.

1. Techniques

Many different techniques are used when compressing images. A selection is described in short, to give an impression of how compression works:

- Reduction of color depth is a simple way to achieve compression, for example storing an image with a color palette of 256 colors reduces the size of a true color image nearly to one third. This is a simple form of scalar vector quantization.
- Conversion to a color system that uses full luminance data and reduces the chrominance data. This works because the human eye is not as sensitive to color information as it is to luminance.
- Run Length Encoding (RLE) stores the number of pixels of one color and the color instead of storing the color for every pixel. Not suitable for photographic images, but may help with computer generated slides and screen shots.
- Difference encoding: Instead of storing the value for one bit the difference to the value of the neighbor is stored using fewer bits. Applied in many compression algorithms.
- Prediction: The value of the neighbors is used to predict the value of a pixel. Stored is usually the difference to the predicted value. Very useful in lossless compression.
- Filtering: Some manipulation is done with the data to allow for better compression with another technique, e.g. using difference encoding and prediction techniques
- Lempel-Ziv-Welch: Works similar to RLE, but uses patterns together with a lookup table instead of bytes [6], Works very well for text.
- Entropy encoding: This is closely related to the Huffman code [7]. The most frequent used terms of a given alphabet will be assigned the shortest codes. Arithmetic coding is an improvement of Huffman coding.
- Pyramids: The image (or a transformation of this image) is decomposed the in a tree like manner. Nodes will only be further decomposed until they consist of a single value.
- Discrete Cosine Transform(ation) (DCT): Is a relative of the Fourier transform and likewise gives a frequency map. This technique is lossy due to roundoff errors.
- Fractal: A system of iterated functions is selected for the description of the image. Offers good compression; scaling is possible.
- Discrete Wavelet Transform (DWT): Uses Haar-functions to code images. Is one of the most promising techniques today. Our experiments have shown better compression rations compared to DCT because the distortions look more natural in current implementations.

1. File Formats

Different frequently used file formats are introduced:

- PCD: Photo Compact Disc. Kodak. Used for storage of images in the image pac file format on a Photo CD. Different resolutions from 128*192*24 to 2048*3072*24 are stored. The YCC color space is used, which is compatible to the CIE-Standard. Data is compressed about 1:6, some chrominance data is lost during compression, however invisible.
- TIFF: Tag(ged) Image File Format. Aldus, Microsoft. TIFF is a catalog format, which includes different compression schemes and image types. The latest available version is v6.0 however only few TIFF images use the v6.0 specification as it is very complex. Due to this, different subsets of TIFF have been implemented, which leads to problems with incompatibility.
- PNM: Portable Anymap file format. Widespread used in the UNIX world. There are many tools available for manipulation and conversion in many different formats. PNM tools operate on portable bitmaps (PBM), greymaps (PGM), and pixmaps (PPM). They can be easily used on a UNIX based Web server to provide an image as requested by the user, changing format, size, gamma, etc. on the fly.
- FIF: Fractal Image File format. Iterated Systems. After starting with ideas of 10,000:1 compression, FIF has developed to „normal" file format. As a rule of thumb JFIF is better up to a compression ratio of 1:40, FIF at higher compression ratios. With FIF the quality of the compression result can be chosen, affecting the file size and/or computing requirements. One advantage of FIF is that images are scaleable and distortions in images look more natural to humans. FIF compression is very demanding regarding computing power. Due to this special compression hardware is needed for compression, decompression is possible using software only.
- JFIF: Joint Photographic Experts Group (JPEG) Image File Format [4,8,9]. The International Standards Organization (ISO) has standardized only the JPEG compression but not the file format. At C-Cube Microsystems two JPEG-based file formats have been defined: JFIF, the de-facto

standard on Usenet, and TIFF 6.0 (TIFF/JPEG) which is supported only by few programs. There may be problems loading JPEG images from a Macintosh because they are sometimes wrapped inside the Mac-specific PICT structure. JPEG is based on discrete cosine transform and spread widely. It is very fast, convenient and the user can use several options, for example vary the image quality, which influences the file size. Due to this it is the major file format used in the Internet for photographic color and greyscale images. A compression factor between 10 and 50 may be achieved, depending on the image for no visual difference to a human observer. It is not very well suitable for computer generated images like presentation slides or screen shots. In this case GIF, TIFF or PNG would be a better choice. The progressive version was introduced for network usage, however, is not yet supported widely.

- GIF: Graphics Interchange Format. CompuServe. Two versions: GIF 87a and (extension of the GIF87a). GIF uses lossless compression (LZW) and allows for up to 256 colors using palettes. Unfortunately, CompuServe has the LZW patent and announced to enforce it (http://rom.oit.gatech.edu/~willday/gif.html). This was one reason for starting PNG. GIF is widespread used in the Internet as it offers good compression, color palettes, and some other nice features like transparent colors. However, it has also some problems: It does not provide more than 256 colors, and does not care about gamma correction.
- PNG: Portable Network Graphics [10,11]. PNG was started as the problems with the LZW patent began. It retains and improves important features of GIF including: Palette-mapped images, streamability, progressive display, transparency, ancillary information, complete hardware and platform independence, effective and lossless compression. In addition new features have been added: Support for true color images of up to 48 bpp, greyscale images of up to 16 bpp, full alpha channel (general transparency masks), gamma indication, detection of file corruption, faster initial presentation in progressive display mode. The format is simple (!), portable, interchangeable, and extensible without loosing interchangeability. The compression with PNG is clearly superior to that achieved with TIFF. We tested 100 of our images and got a compression ratio of 2.23 with PNG but only 1.29 with TIFF. Although PNG is much more suitable for Web usage than GIF, PNG is only supported by a few browsers.
- WVL: Wavelet. This is just one of the many wavelet formats available (e.g. [12]). Wavelet compression is slowly starting to appear in commercial programs, for example Corel Photopaint (enclosed with Corel Draw 6). It is offering good compression especially using high compression factors, because the distortions look much more natural compared to the blocking of JPEG. However, the implementations available right now are not compatible. Thus, if you are setting up a big image collection and do want to use wavelet compression keep a lossless compressed image in an archive.

A good starting point to learn about file formats is the Fileformats FAQ [13].

1. Plug-Ins and viewers

In the beginning of the Web the user was limited to GIF as graphics file format, later JPEG appeared for inline images. Additional formats were available using viewers only. Viewers worked fine, especially considering that they were specialized for a specific application. JPEG images, for example, look much better if decoded in a viewer instead of the browser. This is due to the slower but more precise calculation.

With plug-ins [14] it is now easier to use the image formats optimal for the application. Plug-Ins exist for wavelet compression, fractal compression, vector images, etc. The advantage over viewers is the direct integration in the browser. Thus, depending on the user community, wavelet or fractal compression may now be a better alternative to JPEG, GIF, or PNG. Further, plug-ins often provide additional functions like zooming. However, it has been shown, that only a few users are willing to download plug-ins for viewing images. This may be a problem unless you have a closed group accessing your information.

1. Database

Perhaps we should start with a definition, what is an „image database"? A strict definition would be:

An *image database* is a database that stores images and supports the datatype image with a set of functions.

These functions should include conversion from and to different file formats, scaling, change of color depth, tiling, regions of interest, content based retrieval (CBR), filtering, etc. However, according to this definition, to date there are hardly any image database systems available on the market. People tend to talk of an image database when the database is able to store binary large objects (BLOBs), sometimes viewing capabilities are offered. Because hardly any database provides appropriate functions for image management, it is quite common to store images in the file system and use the database only for links and administrative data.

1. Database Types

There are different kinds of database management systems (DBMS) available:

- Network
- Hierarchical
- Relational
- Object oriented
- Text based

Network and hierarchical database management systems are still holding the most data stored in databases. However in the context of image databases and WWW they can be ignored anyway. There are many more terms containing database, for example deductive databases, active databases, scientific databases, real-time databases, time databases, etc. These do not refer to a DBMS itself, but to special features required for specific applications, which usually are implemented using relational or object oriented technology.

1. Relational DBMS

Relational database management systems (RDBMS) could be called state of the art looking at the usage numbers, the standardization of query languages and available products. The idea for RDBMS was born in 1970 with the relational model by E.F. Codd. „A relational database allows the definition of data structures, storage and retrieval operations, and integrity constraints. In such a database, the data and relations between them are organized in tables. A table is a collection of records and each record in a table contains the same fields. Certain fields may be designated as keys, which means that searches for specific values of that field will use indexing to speed them up. Records in different tables may be linked if they have the same value in one particular field in each table." [15] RDBMS have been designed for business applications like accounting. The idea was to hold the data separate from the application program in a reliable manner. They are flexible and well known and provide the standardized Structured Query Language (SQL) for access to the data. RDBMS are not very well able to deal with multimedia objects and very complex structures. Links between objects are only possible using the equality of keys. Multimedia data types are usually only supported with binary large objects (BLOBs) and hardly any functions.

1. Object Oriented DBMS

Object oriented database management systems (OODBMS) have hardly any market share today (1%<) and are prognosed to have still hardly any impact in the year 2000 (2%). However, they have some advantages for multimedia applications due to the object oriented paradigm, which allows to define methods for classes (data types) and inheritance, which allows to create new data types based on existing ones. „The relationship between similar objects is preserved (inheritance) as are references between objects. Queries can be faster because joins are often not needed (as in a relational database). This is because an object can be retrieved directly without a search, by following its object id. The same programming language can be used for both data definition and data manipulation. The full power of the database programming language's type system can be used to model data structures and the relationship between the different data items. Multimedia applications are facilitated because the class methods associated with the data are responsible for its correct interpretation" [15]. OODBs allow to directly integrate DB access into the code, e.g. to create a record you only have to declare the object type as persistent and create a new instance, the rest will be done by the OODBMS. As methods are related to the objects these may be stored in the database, too. So the strict separation between data and functions as proposed by RDBMS is blurred. Because OODBMS allow other than elementary data types, SQL cannot be the query language. Therefore the Object Query Language (OQL) has been standardized by the Object Database Management Group (ODMG).

1. Text Oriented DBMS

Text oriented database management systems or information retrieval systems were designed to get meaningful results when searching in huge amounts of documents, e.g. literature search. They often allow for natural language querying. To get meaningful results documents are indexed when stored in the database. For relevant words so called „inverted lists" are maintained which tell which documents contain them and how important they are, e.g. „computer" in technical context provides hardly any information, however in medical context it is really significant. When querying, these lists are used to calculate a score for documents. Documents with the highest scores will be shown as hits. A very common techniques which leads to good results is relevance feedback where the user copies passages from a good result and uses this as starting point for the next query.

1. Image Description

As mentioned above, content based retrieval (CBR) to date can only be done for very simple applications. This implies, that additional description is necessary for the images. They could be described using plain text or in a more structured way using key systems or even conceptual graphs.

1. Key Systems

Key systems are widely used to define keywords for retrieval. They limit the vocabulary and thus enable a more unified description. With key systems it is easy to deal with synonyms and multiple languages. Key systems can be chosen hierarchically, that means that each node has exactly one parent node (except the root). With this construction different kinds of queries can be asked. This can be demonstrated with an anatomy example. A search for images with localization 'finger' could lead to different results depending on the mode:

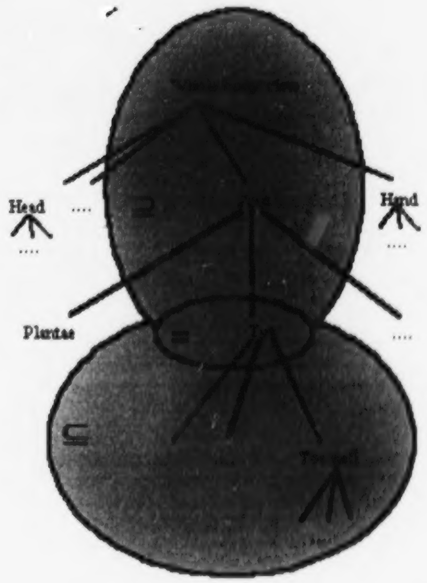


Figure 1: Search using a hierarchical key

'=': All images with exactly the localization 'finger' will be retrieved

'': All images with 'finger' or parts of the 'finger' will be retrieved

'': All images in which a 'finger' is visible will be retrieved

There are a lot of (hierarchical) key systems available, e.g. the International Classification of Diseases and Causes of Death (ICD), Medical Subject Headings (MeSH), Systematized Nomenclature of Medicine (SNOMED), and lots, lots more. In every application area there may be specific commonly used key systems. However, the real world does not always fit well into tree like key systems, even if multidimensional systems are used.

1. Conceptual Graphs

BEST COPY AVAILABLE

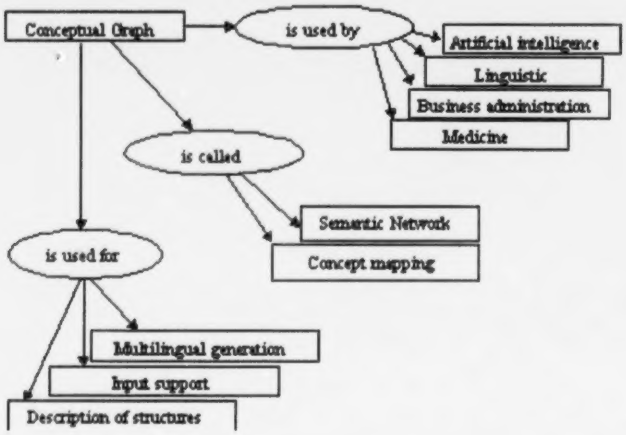


Figure 2: Description of conceptual graphs using a conceptual graph.

In fact the world can be much better represented as a network (or should I call it web:-). That is the starting point for conceptual graphs. Conceptual graphs [16-19] are a highly readable graphical language, with a great expressive power and can be easily mapped to and from natural language. They are an intuitive knowledge representation language that graphically depicts relationships between concepts. A formalized description of the concept space, that contains concepts and relations, requires computer support in real applications because the complexity is big. Once a concept space has been developed computers are able to check the semantics of a statement and may reject statements that are not correct. This can be done e.g. by offering only terms for description which are allowed in the current context. A simple editor has been developed at the University of Hildesheim, the Concept Space Editor (CEPTER) [20], a more comprehensive system is developed in the Peirce project [17-19].

One very complex application of conceptual graphs is the Unified Medical Language System (UMLS) [21] of the National Library of Medicine (NLM). The so called Semantic Network contains 133 semantic types and 50 relations which link pairs of semantic types.

1. WWW Interface

When creating a (distributed) image database there are several choices for the front-end, which is not necessarily the WWW. If you have got a really distributed database sometime, you may be better off, not using the WWW. However, there are many good reasons considering the WWW as front-end, which in turn may be not the only one.

The WWW gives much functionality for free, including transparent network access, decompression of images, gradually display of images, hypertext/-media, security mechanisms, and being able to reference the whole WWW. In the WWW many people have been thinking about data types and developing viewers even for such data types as virtual reality scenes. The Web uses the Hypertext Markup Language (HTML) which is somehow standardized and thus guarantees the usability of the generated resources for a long period of time. You do not have to care about different platforms. Even if you write code with ANSI C in a very portable manner there will be platforms which do not like this code. With the WWW you are fine off, there are companies caring for these problems. The data you offer is available all over the world. If you like, you can restrict the access to the users you like to be able to access your data. Usually you will need only a single user database license. The graphical user interface is easy to use and well known, people do not have to deal with different platforms and keyboard codes. Last not least the WWW is developing very fast, new tools are announced nearly every day. These tools can be used e.g. for your database perhaps allowing to provide „database access" to your data without even having a database.

BEST COPY AVAILABLE

Unfortunately there are also some limitations of the WWW:

- There are only few interface elements and control of the representation is limited.
- Every action you take usually will require an interaction with the server.
- A stateless protocol may have some problems for your application. However at the first WWW conference it was reported [22] that even this may be an advantage as it forces a different, very modular design.

1. File System versus Database

In the WWW traditionally the file system („server tree") has been used to provide documents. There are, however, good reasons to use a database instead of a file system for special applications or even for the whole server tree.

A database allows comfortable maintenance of the data, for example:

- Administrative data can be easily included into documents
- Unified appearance of all documents on the server can be ensured
- All documents of a specific author can be retrieved
- User profiles can be used to offer multiple languages
- Filters can translate document types in the requested form
- A database has been constructed to store data in a reliable way, that is, each database has to comply to the Atomicity, Consistency, Isolation, Durability (ACID) principle of transactions. These features are important if data in the database should be changed.

As advantage of the file system there are two points I want to mention:

- It is really easy to use, as everybody has to use it every day
- There is only one system to care for. There are no problems with another, perhaps remote system, the database (-server). I think you all know these troubles with different operating systems, versions, patches, etc. :-)

Using a file system you can offer static and dynamic connections. A *static* connection is one which uses only the server tree. This allows fast access, can be used by robots and the tree structure can be easily understood by the user in the most cases.

We use the term *dynamic* connection if a Common Gateway Interface (CGI) script is involved. These kind of scripts will usually have names like getpage.pl and will be called with parameters including user specific. Such an URL could look like this:

`„http://my.server/cgi-bin/getpage.pl?document=doia.htm&lang=de&user=bittorf"`

In this case the document doia.htm will be fetched in the language „de" (German) considering the user profile of „bittorf". This user profile may contain information about the allowed IP addresses, the display resolution of the associated computers, gamma correction parameters, etc. The document may then be processed according to this user profile before being offered to the user. The drawback of dynamic connections is that there is a (much) bigger server load due to preprocessing.

1. General Architecture

For a basic WWW database interface the following steps are necessary:

1. Generate the user interface (HTML form)
2. Receive user input via HTTP (e.g. WWW server with an Common Gateway Interface (CGI) or application programming interface (API))

3. Analyze user input, derive query, and pass it to the DB client
4. Access the database (e.g. network client for a database or a server demon process)
5. Generate HTML from the output and send it back to the user

---

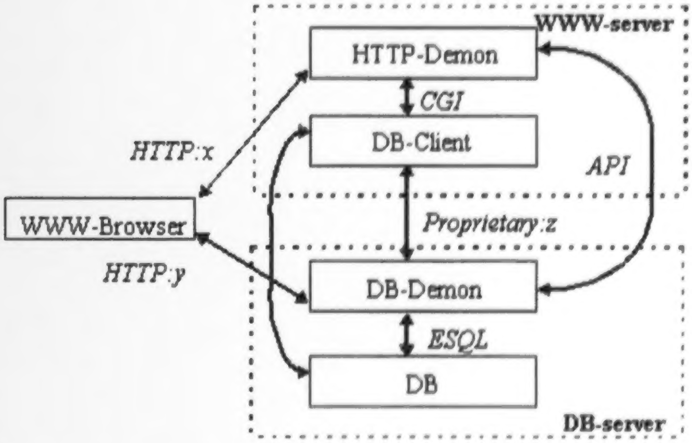


Figure 3: Different ways of database integration

Figure 3 shows different ways to achieve this. The easiest way is using a static HTML form, processing it with a CGI script, and using a network client of the database. Using the application program interface (API) of the webserver instead of the CGI speeds up the processing, but is less portable.

A little bit more sophisticated and faster is a database demon process that keeps the database open and waits for queries instead of the network client. Depending on the speed of the network connection and the size of the database, this may speed up queries by a factor of five or so. Another improvement is a HTTP interface to the demon process thus offering a database that can be directly accessed from the browser. Our own experiments have shown this may again speed up queries by a factor of five or so, depending on the speed of the network connection to the Database-/Webserver.

1. Forms interface

The forms interface allows the user to formulate a query or other database operations using the following HTML 2.0 elements [23]:

- Input Field: INPUT
    - Text Field: INPUT TYPE=TEXT
      Allows for text input.
      ```
      <input name=zip size=10 maxlength=10 value="99999-9999">
      ```
    - Password Field: INPUT TYPE=PASSWORD
      As above, except that the value is obscured as it is entered.
      ```
      <input type=password name=passwd>
      ```
    - Check Box: INPUT TYPE=CHECKBOX
      Select some of the choices.
      ```
      <input type=checkbox name=flav value=vanil checked>Vanilla<br>
      <input type=checkbox name=flav value=straw>Strawberry<br>
      <input type=checkbox name=flav value=choco checked>Chocolate
      ```
    - Radio Button: INPUT TYPE=RADIO
      Select one choice.

```
<input type=radio name=flav value=vanil>Vanilla<br>
<input type=radio name=flav value=straw>Strawberry<br>
<input type=radio name=flav value=choco checked>Chocolate
```

  - Image Pixel: INPUT TYPE=IMAGE
    Image maps. Clicking implies „submit".
    ```
    <input type=image name=point src="map.gif">
    ```
  - Hidden Field: INPUT TYPE=HIDDEN
    No user interaction. Useful e.g. to transmit information about the session or user profiles.
    ```
    <input type=hidden name=user value="bittorf">
    ```
- Selection: SELECT + Option: OPTION
  Allows user to select one or more elements of a given list.
  ```
  <SELECT SIZE=2 MULTIPLE NAME="flavor">
  <OPTION>Vanilla
  <OPTION>Strawberry
  <OPTION value="RumRasin">Rum and Raisin
  <OPTION selected>Peach and Orange
  </SELECT>
  ```
- Text Area: TEXTAREA
  Like „Input Field", except that there are more columns and rows possible. A multi line initial value may be given to the textarea.
  ```
  <TEXTAREA NAME="address" ROWS=6 COLS=25>
  Andreas Bittorf
  Hartmannstr. 14
  91054 Erlangen
  </TEXTAREA>
  ```

All information entered into this form plus some extra information will be available in the CGI script processing the query, which is given in the FORM ACTION field.

HTML 3.0 suggests new elements [23]:

- Client side scripts for range checking (SCRIPT)
- Range controls (sliders, or knobs)
  ```
  <input name=rating type=range min=1 max=10>
  ```
- Single/multiple choice menus
- Scribble on image (type=scribble)
- File widgets for attaching files to forms
  The ACCEPT attribute can be used to specify a comma separated list of MIME content types. These are used to restrict the kinds of files that can be attached to the form.
  ```
  <input name=pictures type=file accept="image/*">
  ```

HTML 3.0 will never become a standard, (because of the slow process of standardization?!) and has been replaced by the HTML 3.2 specification that incorporates some of the suggestions of Netscape (Navigator) and Microsoft (Internet Explorer). Important features like FIG or FRAME have not yet been "standardized", but tables and applets are now available. (The HTML tables specification (RFC1942) has been ratified as an IETF proposed standard.) Thus, for widest portability I suggest to use HTML 2.0 elements and tables in addition.

If you want to use frames please provide a version that is also usable without a browser able to display frames. Applets should also be used careful, because there have been security problems in the past and thus many users may not be willing or allowed to enable applets. Keep in mind: enabling an applet means that a (perhaps malicious) program is executed at your computer.

How the further standardization process of HTML will continue is not clear.

1. The Common Gateway Interface

After a form has been submitted the WWW server calls the corresponding CGI program with the

BEST COPY AVAILABLE

variables as parameters. In addition to the user input, server variables can be accessed. Unfortunately, different servers offer different variables. Important ones that are used by the most servers are for example: HTTP_USER_AGENT (which client), HTTP_ACCEPT (MIME formats that the client understands), REMOTE_ADDR (IP-address of client), REMOTE_HOST (name of client), GATEWAY_INTERFACE (kind, version). Other variables depend on the client that is used, for example HTTP_UA_PIXELS, and HTTP_UA_COLOR can be used with Mircosoft Internet Explorer 3.0, and HTTP_REFERER with Netscape Navigator 3.0

In the first step the CGI program has now to decode the variables and to derive a query. Then it can send the query, e.g. to a database server demon.

Note: Be very careful when writing or using CGI scripts, especially if you are doing it in PERL. There are a lot of security problems. Read the World Wide Web Security FAQ [24] to get an idea.

1. Database server demon

Again, because it is really important I want to mention the database server demon. It should open the database and keep it open to avoid the time needed to open the database for every request [25]. After the demon has opened the DB, it should provide a socket where requests can be sent. Queue and socket management with some kind of dispatcher are a nice additional features. The output of the query can be returned to the CGI/API script via the socket. The script then formats the HTML output depending on the user profile and the results.

1. Tools

A lot of tools for interfacing a database with the WWW have been developed by users and database vendors. O2 offers O2Web, Oracle the Oracle WWW Interface Kit, IBM DB2WWW, and there are many, many others. For Windows and Windows NT there are a lot of programs available allowing the usage of the open database connectivity (ODBC) interface and high level languages like visual basic. There are commercial as well as freeware products. There are big differences concerning performance and functionality. Some interfaces are used only for querying, others also support extended operations like updates, inserts, or even create table operations. A good overview of existing products is given at http://cscsun1.larc.nasa.gov/~beowulf/db/existing_products.html. Further the collection at http://www.stars.com/Vlib/Providers/Database.html is recommended. Of course, we cannot discuss the different solutions here in detail.

To give you a first impression two simple approaches are discussed. GSQL was one of the first (1993!) WWW-DB interfaces that was developed. It is based on the CGI and is not very comfortable, but gives an idea how such an interface works. The advantage of the CGI solution is the portability. Another simple but vendor specific solution is the mechanism provided by the (free) Internet Information Server of Microsoft for MS Windows NT Server 3.51/4.0. It uses some kind of server API (dynamic link libraries) and allows to access any ODBC database.

1. GSQL

Generic SQL (GSQL) [26] supports only querying a database. It provides a form description with a so called proc file. In this proc file the appearance of the form for the user input and the actions to be taken at submit are described. For Sybase as well as other databases a corresponding backend is available. GSQL does two jobs:

1. It generates a form for the user input.
2. From the input it generates a SQL string according to the given rules and forwards it.

The CGI script just calls gsql with the proc file as parameter:

```
#!/bin/sh
```

```
/gsql/proc/db_query $QUERY_STRING
```

In the following proc file the user interface and construction of a SQL string are described. There is plain HTML which can be offered through the „TEXT" element and there are Form elements which are related to the SQL string via the WHERELIST and SELECTLIST. All elements are explained in the proc file in more detail. GSQL keywords have been written with bold letters. Figure 4 shows how the Forms interface GSQL generates of this proc file looks like.

```
# -------------------------------------------
# db_query:
#
# GSQL proc-file as front end for DOIA
#
# This version uses a database view
# -------------------------------------------

#TEXT: just display the text (may include HTML-Tags!)

TEXT <title>DOIA: Advanced Query Interface to the SQL-Database</title>;

#The "Heading"-line will be repeated when displaying the result.
It is quite long as it contains the navigation icons and the heading

HEADING <a href="http://www.uni-erlangen.de/docs/derma/"><img align=bottom height=39 width=40
src="http://www.uni-erlangen.de/docs/derma/home_k.gif" ALT="Home"></a> <a
href="http://www.uni-erlangen.de/docs/ derma/bilddb/db.htm"><img align=bottom height=40 width=66
src="http://www.uni-erlangen.de/docs/derma/bilddb/bilddb_k.gif" ALT="Image Atlas"></a> <a
href="http://www.uni-erlangen.de/docs/ derma/personen/mailandi.htm"><img align=bottom
src="http://www.uni-erlangen.de/pict/images/comm/email.gif" ALT="Feedback"></a><hr><H1>DOIA: Query
Interface to the SQL-Database</H1>;

TEXT Queries are case sensitive.;

TEXT Please report problems (please note that the DB-server may be down sometime).<p>;

#Advanced HTML-tags can be included

TEXT <TABLE border=5>;

#The variable following "SUB" is included in the "WHERELIST", that is it will become part of the
constructed SQL-statement, "AS" bild.bi_id >= the value of this variable

SUB s_bi_id_v WHERELIST AS bild.bi_id >= $;

#The value of the variable will be entered in a simple textinput field ("FIELD"). As Text ("TITLE")
"Image ID >=" will be displayed. The HTML-tags belong to the table.

SHOW s_bi_id_v TITLE "<tr><th>Image ID<td>&gt;=" FIELD;

SUB s_bi_id_b WHERELIST AS bild.bi_id <= $;

SHOW s_bi_id_b TITLE "<br>&lt;=" FIELD;

SUB s_cd_v WHERELIST AS bild.bi_cd_nr >= $;

SHOW s_cd_v TITLE "<th>CD-No<td>&gt;=" FIELD;

SUB s_cd_b WHERELIST AS bild.bi_cd_nr <= $;

SHOW s_cd_b TITLE "<br>&lt;=" FIELD;

SUB s_a_v WHERELIST AS bild.bi_alter >= $;
```

BEST COPY AVAILABLE

```
SHOW s_a_v TITLE "<tr><th>Age<td>&gt;=" FIELD;

SUB s_a_b WHERELIST AS bild.bi_alter <= $;

SHOW s_a_b TITLE "<br>&lt;=" FIELD;

SUB s_sex WHERELIST AS bild.bi_geschlecht = '$';

SHOW s_sex TITLE "<th>Gender (M/W for male/femal)<td>" FIELD;

#Depending on the SQL of the database-backend more complex operators can be used, e.g. substring match
with like and wildcards. If our database would support a character conversion to upper case (Ucase), we
could offer a case insensitve conversion. Stored procedures in the database can thus offer a lot of
functionality to the Web-frontend.

SUB s_diag_s WHERELIST AS bild.di_text like '%$%';
```

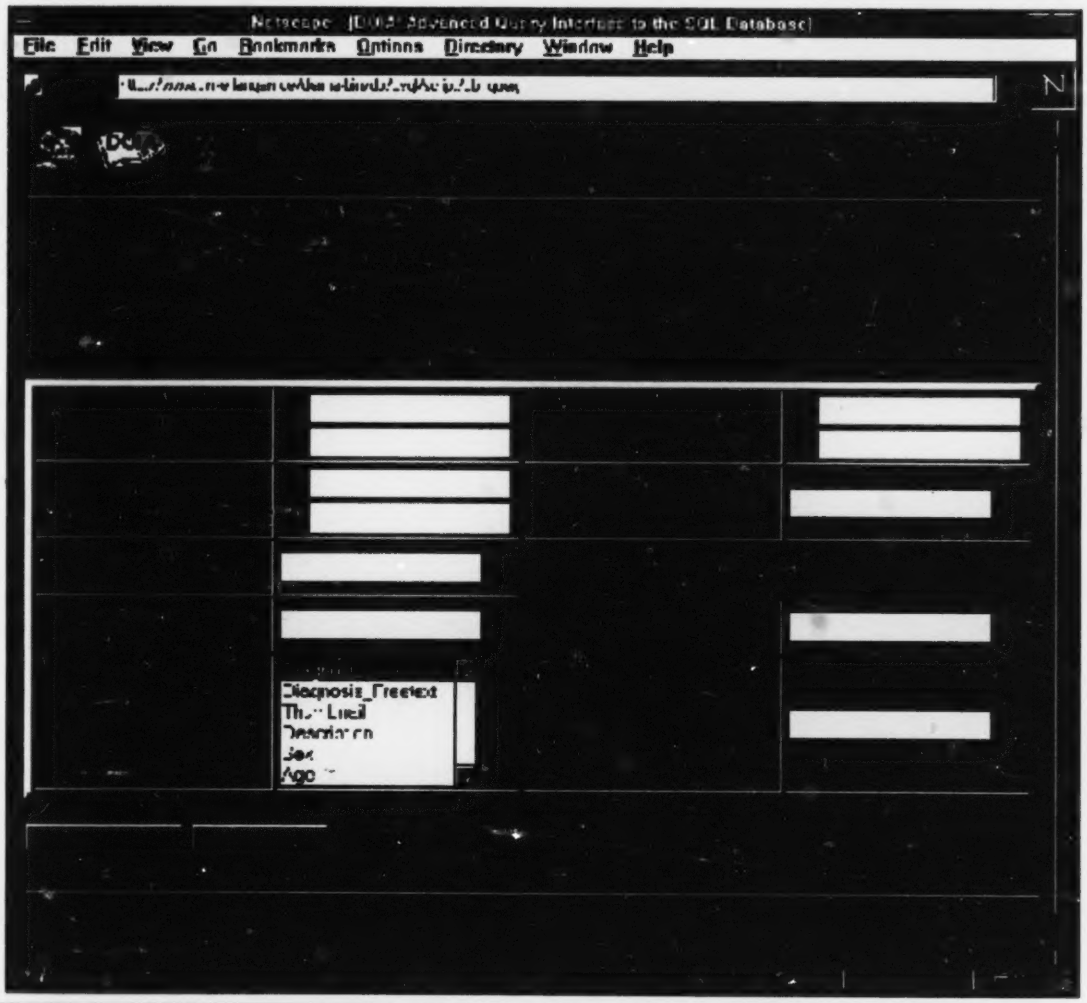

Netscape - [DOQA: Advanced Query Interface to the SQL Database]
File Edit View Go Bookmarks Options Directory Window Help
N
Diagnosis_Freetext
Description
Sex
Age

Figure 4: Input interface corresponding to GSQL-Proc file

```
SHOW s_diag_s TITLE "<tr><th>Standardized Diagnosis (substring)<td>" FIELD;

SUB s_diag WHERELIST AS bild.bi_diag_text like '%$%';

SHOW s_diag TITLE "<tr><th>Diagnosis text (substring)<td>" FIELD;

SUB s_desc WHERELIST AS bild.bi_besch_text like '%$%';

SHOW s_desc TITLE "<th>Description<br>(substring)<td>" FIELD;

TEXT <tr>;

#Similar to the "WHERELIST" the "SELECTLIST" contains additional fields to be displayed. The second "as"
is part of the SQL-String to be generated: "select a as Name, b as Date from c". The Diagnosis after the
sub is again the variable name.

SUB Diagnosis SELECTLIST AS bild.di_text as Diagnosis;

SUB Diagnosis_Freetext SELECTLIST AS bild.bi_diag_text as Diagnosis_Freetext;

SUB Thumbnail SELECTLIST AS bild.bi_klbild_ref as Thumbnail;

SUB Sex SELECTLIST AS bild.bi_geschlecht as Sex;

SUB Description SELECTLIST AS bild.bi_besch_text as Description;

SUB Age SELECTLIST AS bild.bi_alter as Age;

# A variable beginning with a _ means an indirection (-> variables like Thumbnail etc. will be used
instead) Using this mechanism the display and semantic can be separated.

SUB _displaylist SELECTLIST AS $;

#The variable _display list will be shown using a scrollbox "SCROLL"

SHOW _displaylist TITLE "<th>Select<br>for display<td>" SCROLL Diagnosis, Diagnosis_Freetext, Thumbnail,
Description, Sex, Age;

SUB s_diag_o WHERELIST AS bild.di_nr_ordnung = $;

SHOW s_diag_o TITLE "<th>Avoid synonyms for diagnoses by entering 1<td>" FIELD;

#The "SQLPROG" contains the program which accepts the SQL-statement, computes and displays the result.

SQLPROG /home/www/www/derma-bin/db/gsql/sqlperl.pl;

#In the "SELECTLIST" all fields are mentioned which will be displayed in any case. Again SQL keywords
(here: "Distinct") may be used.

SELECTLIST DISTINCT bi_id AS Ident , bi_html_ref AS Large_Image;

#The "FROMLIST" contains the FROM part of the SQL statement. We take advantage of the fact that in SQL
aliases can be used, that is we use the view bild_diag instead of the table bild, but we did not have to
change any other code in the GSQL script. Of course more than one table can be used.

FROMLIST bild_diag bild;

#With the "WHERELIST" restrictions can be applied, e.g. for joining two tables. We did not use this
feature as we preferred to create a suitable view on the database. We have given two examples, the first
restricts a field and the second joins the tables bild and bild_url.

#WHERELIST bild.di_nr_ordnung = 1;
```

```
#WHERELIST bild.bi_id = bild_url.bi_id;

TEXT </TABLE>;

#For other systems it may be necessary to define some variables

#DEFINE LOGIN something;

#DEFINE PASSWORD whatever;

#DEFINE SYBASE_ENV SYBASE=/usr/local/sybase;
```

If you would like to modify the appearance of the document (as we did), e.g. submit button or title, you have to have a look at the gsql.c source.

1. Internet Information Server

With the Internet Information Server (IIS) all databases accessible via the Open Database Connectivity (ODBC) interface can be used. It uses two kind of files to define the interface:

- Internet database connection (*.idc) files
- HTML extension (*.htx) files

In contrast to GSQL, where the form is described in the proc file and generated from this file by a script, with IIS the HTML form has to be set up manually for user input. The database can be accessed with the IDC-file as "form action" element, like in the following simple example:

```
<FORM action="/scripts/samples/bilddb.idc" METHOD=get>

CD-No <INPUT NAME="CdNr" VALUE="" size=10><br>

Img-No <SELECT NAME=RelCdBiNr>

<OPTION> =

<OPTION> <=

<OPTION> >=

</SELECT>

<INPUT NAME="CdBiNr" VALUE="" SIZE=10>

<P>

<INPUT TYPE="SUBMIT" VALUE="Start query"><INPUT TYPE="RESET" VALUE="Clear Form">

</FORM>
```

The corresponding IDC file ("`/scripts/samples/bilddb.idc`") looks like follows:

```
Datasource: Bild DB

Username: sa

Template: bilddb.htx
```

BEST COPY AVAILABLE

```
RequiredParameters: CdNr

SQLStatement:

+Select *

+FROM Bild

+WHERE Bi_Cd_Nr=%CdNr% AND Bi_Cd_Bild_Nr %RelCdBiNr% %CdBiNr%

DefaultParameters= CdNr=1, CdBiNr=5
```

The "datasource" (Bild DB) has been previously defined via the System (!) DSN and is an Access database. The "template" gives the HTX-file that defines how the result will be processed. The query is only executed if the "RequiredParameters" have been entered in the form. The "SQLStatement" defines how the query looks like, parameters from the input form are included enclosed in '%'. Note, that in this example not only values but also qualifiers ('<=", "=", and ">=") have been used as parameters. Further keywords are available, e.g. "MaxRecords" to limit the size of the result set, "MaxFieldSize" to set the maximum length of data returned per field, and "Password" to set the user connecting to the database.

After processing the SQL statement a HTML file is constructed using the specification in the HTX file. The section for the display of the output could look like:

```
<h1>Query result</h1>

<%begindetail%>

<%if CurrentRecord EQ 0 %>

<h2>Selected images:</h2>

<%endif%>

Data: <a href="/bilder/cd<%Bi_Cd_Nr%>/img000<%Bi_Cd_Bild_Nr%>.htm"> <b>CD <%Bi_Cd_Nr%>,
<%Bi_Cd_Bild_Nr%></b></a>

<%enddetail%>

<%if CurrentRecord EQ 0 %>

<h2>Sorry, no match.</h2>

<%endif%>
```

The section between "`<%begindetail%>`" and "`<%enddetail%>`" is executed once for every row of the result set, that implies that it is not executed if the result set is empty. Variables from the database and the user input can be used as "`<%attrib_name%>` respectively "`<%idc.var_name%>`". Using the "`<%if ..%>`" statement, conditional blocks can be included. Server variables can be used, for example "HTTP_ACCEPT", "HTTP_USER_AGENT", "HTTP_UA_PIXELS", and "HTTP_UA_COLOR". Thus the output can be configured e.g. depending on the capabilities of the client.

1. Database versus Information Retrieval

The Web is often used only as a means to present information. If it is not used to acquire data no updates, inserts and deletes will be necessary through the web. In this case information retrieval components may be more suitable than a database. A database supports transactions and is best suitable

BEST COPY AVAILABLE

for use with exact information and the combination of attributes.

Information retrieval works with fuzzy logic. The user gives some terms, even a natural language query like „I would like to find all papers dealing with images, databases and the WWW" would be appropriate. In the first step the system eliminates useless words in the query and expand the relevant terms using a thesaurus. After this an inverted list is used to find documents containing the significant words and their score. Calculating the score for a word in a document may be done considering the frequency of appearance in the text and the document universe, the length of the paper, and the place of appearance, for example are title, header, and abstract more important than plain text. Knowing the scores for the single key words a combined score will be calculated for the best N hits. Doing information retrieval often leads to much better results than using a database: Have you ever tried to use KnowledgeFinder as front-end to Medline?

So, use a database if you:

- would like to do updates, inserts and deletes in a reliable way
- usually use exact match, for example the patients name and date of birth
- would like to combine different pieces of data

Use information retrieval if you:

- need to search in text based information
- have no clear understanding of possible terms and values
- would like to have a ranking of appropriateness of search results

You do not have to decide for a database or an information retrieval system, because modern databases support full text search: Illustra uses a "Text DataBlade", Oracle the "Oracle ConText Option", Sybase "Topic", and IBM DB2 the "DB Text Extender".

1. Interface Design

There are a few rules how to design a user interface. The three and most important are: simple, simple, and simple. Have a look at the Web based search engines like Lycos [27], and WebCrawler [28]. They have developed from an input interface with many parameters to a single input box. Alternatives are not given any longer in the entry screen. Good default values are used for the novice user. The experienced user knows where to change the settings, if necessary.

1. Integration of distributed data

With distributed data we face two problems:

1) Transfer times of data can be very long

2) A (at least weak) consistent view on the data is required

1. Access times

To reduce access times there are two mechanisms currently used: mirroring and caching.

Mirroring is quite common with FTP archives. To make optimal use of the mirrors, ARCHIE allows to locate the closest copy of a file. However, the Web cannot directly be compared to FTP because: a) files tend to be much smaller and b) it is not common to use a search machine to locate just one file and most Web users don't even know or obey netiquette. Thus, for best usage the lookup of the closest file must be somehow automated, e.g. by resolving names depending on the location of the user.

Advantages and disadvantages of replication (mirroring) are as follows:

+ Mirroring guarantees fast access, even in bad connected countries

+ The update frequency can (and should) be selected as required

- Where is the master source?

- Is the information up to date?

- Cache servers may be filled with copies of the same information

Caching is very helpful if some information is required by many people using one common proxy/cache server. Thus caching works best if one server is used by a group with common interests.

Advantages and disadvantages of caching are as follows:

+ Transparent to user

+ Fine for teams and frequently requested information

+ Easy to implement

- Delays every request

- Usually used for diverse groups

- Average hit rates of only 30%

- Slow/no access to information not in cache

1. Consistent view

Another problem with distributed data is to offer a consistent view to the user. Distributed databases provide good mechanisms, but can be only used if very close cooperation exists between the involved partners. Usually this will not be the case for organizations which would like to combine their data for access through the Web.

Fortunately, there are mechanisms to allow common search interfaces without a distributed database, think of Lycos which is, in the widest sense, a common query interface for (nearly) all documents on the Web. However, these search engines are not a) focused on user interests, and b) capable of indexing databases because there is the problem of an infinite query space.

The best way to deal with this problem is to generate a dynamic or static HTML-tree from your database. This is especially helpful if your data is structured in a simple way, for example, if you have stored contact information for your employees you may generate a company phone book structured by departments. This will also allow full text indexing, e.g. with WAIS. An update can be easily done, just generate the structure again and delete the old one.

HARVEST [29,30] is a more complicated, but conceptually sound approach to access distributed information. It does not only use full text retrieval but also offers meta tags for structured information. It operates completely different to robots. The idea is not to search Web servers, but to rely on the information given by them using gatherers and brokers. *Gatherers* collect information about the specified documents on a specified server. They do not only index HTML documents but also postscript, winword, rich text format, word perfect, framemaker, etc. For each document, data is stored in the so called Summary Object Interchange Format (SOIF). Brokers get information of gatherers and other brokers according to defined rules. Further, brokers offer the searchable data to users.

1. Case study: DOIA

As diagnosis in Dermatology relies heavily on visual information we decided to set up an image database, a part of which is now offered through the WWW: Dermatology Online Atlas (DOIA) [31-39]. DOIA is intended to be a basis for a distributed hypermedia textbook of dermatology. It offers selected images from our slide archive with about 100.000 slides. Right now it covers basic diagnoses important for student education and an extensive collection of images from specialists in our hospital. DOIA shall become a distributed resource for reference and teaching of outstanding quality. With experts around the globe contributing, it will offer up to date images and expert knowledge.

Some goals will be achieved through the common image atlas. Images will be stored in a standardized manner and a common language of description will be developed. In Dermatology there is already a very clear description language available, however, dermatologists tend to invent new terms for description and the terms used are not really standardized. Giving a terminology together with reference images may help here. Fast access will be available. It is still quite common to search for literature in Medline and order the papers. Rare and country specific diseases will be covered as specialists around the world will be asked to contribute to this resource. Even more important is that images of different skin types (white, black, asian) will be available for reference. Skin diseases tend to look completely different on different skin types and reference images are usually only available for the most common skin type, e.g. white in Germany. An international reference atlas will help in diagnosis.

1. Images

Clinical and histological images were selected from the archive or continuing medical education (CME) units. These were commercially scanned and stored on PCDs, sometime we also used a slide scanner. Using a PC Pentium 100 with a miro 40sv graphics card and a miro 2085E monitor plus miro proof for calibration, the images were enhanced and made anonymous. After manipulation the images and corresponding thumbnails were stored on a magneto-optical disc and JPEG compressed on hard disc.

Studies [40-43] with clinical photographs (skin, body overview, etc.) have shown that the resolution of 768x512x24 (1.125MB) is sufficient to see every required detail. Further, it displays well on most screens, because usual monitor resolutions currently range from 800x600 to 1600x1200. Another study [40] has shown that images at this resolution can be compressed at about 1:37 with JPEG without loosing diagnostic relevant details. In about 90% of the images not even any difference could be detected between the original and the compressed one.

1. The local database

Using standardized and additionally developed hierarchical key systems the images were described. Data input was done using a Visual Basic front-end to the database. The description is stored in a database, whereas the image remains in the file system, because MS Access, which we use for our local database, had some trouble with storing images. In addition, in the file system images can be used by the database as well as in local HTML documents.

1. Mapping the local Database to the Web

We offer different ways to access our database through the Web:

1. We generate a static access structure to the images using diagnoses (alphabet, ICD).
2. Additional tags are stored as meta information for HARVEST and robots. Full text retrieval is offered using HARVEST with GLIMPSE and WAIS. This query interface suffices for most applications.
3. A gateway to an INFORMIX database is offered that has been designed using GSQL, a local client and a remote database demon. The local client passes SQL to the database demon which returns the results as a set of rows. For the HTML output we use the powerful text manipulation

features of PERL. Updates of the INFORMIX database are done by FTP and the load utility of INFORMIX with a subset of our local database. The database demon can be also contacted directly by the browser via HTTP, e.g. if queries are included as links into documents.

User comments for image documents are supported. These comments are stored in the database, a log file, and the diagnosis access structure. For updates of the local database the log file is used. For retrieval of images stored in Erlangen, the database solution is very powerful, however, an architecture based on the HARVEST system provides some text based search ability and allows to build a distributed image „database" for Dermatology easily.

1. Acceptance and user profile

Usage statistics and personal communication show that the image atlas is used enthusiastically. In the end of 1994 we had about 200 document requests per day, now (8/96) there are more than 10.000, a big number considering that very specialized information is given which is only of interest to dermatologists. Surprisingly, about 30% of the requests came from America despite the slow connection.

In this summer we conducted a user survey. In about two months we got 200 answers. The result is quite interesting because it gives some insight about the users of our server. A short summary of the results:

- International access is quite common only about 50% of the users came from Europe
- 83% of our users use a PC, 13% a Mac, and 4% a workstation
- 80% of the browsers were Netscape Navigator, 7% of Microsoft Internet Explorer
- Over ten different versions of Netscape Navigator were used from different platforms going back to version 0.9b
- Over 75% of the users found our offer online, and more than half of these used search engines
- Access speed: 10% with 9,6 Kbit/s or less; 16% with 14.4/19.6 Kbit/s; 30% with at least 28.8 Kbit/s; 18% with 64 Kbit/s; remaining2 Mbit/s
- ISDN connections were mainly used from Germany and hardly in the US
- Over two third accessed the Web from home
- 50% of our users were physicians, additional 20% had a job related to medicine
- The database was mainly accessed for the purpose it was designed for (multiple selections possible): 50% continuing medical education, 30% reference, 25% teaching, 28% patients
- Hyperlinked access (alphabet 80%, ICD: 18%) was much more used than full text retrieval (22%) or the database interface (15%)
- The overall rating of the atlas and the image quality was very good

1. Summary

You should now be able to answer the following questions

- How should I acquire digital images for my application?
- What compression technique will be appropriate for my application?
- What image file format will be appropriate for my application?
- Is color/gamma correction necessary? What could be done for this?
- What kind of database system should be used?
- How can image information be described for retrieval?
- Do I need a database? Wouldn't a text based system be the better choice?
- Should tools for a Web based database be used?

and to set up your image database in the Web.

Good Luck!

1. Acknowledgments

I want to thank all those who helped me in the DOIA project (in alphabetically order):

- Bergmann, Martin, MSc
- Burzler, Frank, MSc
- Diepgen, Thomas L., MD, PhD
- Fartasch, Manige, MD, PhD
- Hiller, Dorothea, MD
- Holleczek, Peter, PhD
- Killig, Sven
- Krejci-Papa, Niels, MD
- März, Alexander (Student)
- Proschek, Till (Student)
- Rothenanger, Jürgen, MSc
- Schuler, Gerold, MD, PhD

And other physicians of our hospital

1. References

1. Gottheimer D. Imaging und das KODAK Photo CD System. Kodak. 1993; Kodak.

2. Krestel E. Bildgebende Systeme für die medizinische Diagnostik. 2. ed. Berlin, München. Siemens. 1988;

3. DICOM Standard. PS3.1(199x). NEMA. 1996; FTP://ftp.xray.hmc.psu.edu/dicom_docs/.

4. JPEG image compression: Frequently Asked Questions. 1996; http://www.cis.ohio-state.edu/hypertext/faq/usenet/jpeg-faq/top.html or FTP://rtfm.mit.edu/pub/usenet/news.answers/jpeg-faq/.

5. Compression FAQ. 1996; http://www.cis.ohio-state.edu/hypertext/faq/usenet/compression-faq/top.html or FTP://rtfm.mit.edu/pub/usenet/news.answers/compression-faq/.

6. Welch TA. A Technique for High Performance Data Compression. IEEE Computer. 17. 6. 1984;

7. Manstetten D. Schranken für die Redundanz des Huffman-Codes. Hüthig. 1995; Hochschultexte Informatik; 8.

8. ISO/IEC. Information Technology - Coded Representation of Picture and Audio Information - Digital Compression and Coding of Continuous-tone Still Images. 10918. International Standards Organization. 1992;

9. Sloan AD. The fractal image format and JPEG. Miller Freeman Expositions; Boston. 1991; p. 460; Electronic Imaging International.

11. Boutell T, Lane T, et al. PNG (Portable Network Graphics) Specification - Working Draft. Boston. World Wide Web Consortium. 1996; http://www.w3.org/pub/WWW/TR/WD-png.

12. Zandi A, Boliek M, Schwartz EL, Gormish MJ, and Keith A. CREW Lossless/Lossy Medical Image Compression - A contribution for ACR-NEMA MED-PACS Working Group IV. RICOH Clifornia Research Center. 1995;

13. Murray JD. Fileformats FAQ. 1996; http://www.cis.ohio-state.edu/hypertext/faq/usenet/graphics/fileformats-faq/top.html or

FTP://rtfm.mit.edu/pub/usenet/news.answers/graphics/fileformats-faq.

14. Plug-ins. Netscape. 1996; http://home.netscape.com/comprod/products/navigator/version_2.0/plugins/index.html.

15. Howe D. The Free On-line Dictionary of Computing. 1995; http://wombat.doc.ic.ac.uk/.

16. Sowa JF. Conceptual structures. Reading, Mass. u.a. Addison-Wesley. 1984;

17. Proceedings of the 3rd International Workshop on Peirce: A Conceptual Graphs Workbench.Levinson, R. and Ellis, G. editors. 1994 Aug 19; 1994; International Workshops on Peirce.

18. Proceedings of the 2nd International Workshop on Peirce: A Conceptual Graphs Workbench.Ellis, G. and Levinson, R. editors. 1993 Aug 07; Quebec City, Canada. 1993; International Workshops on Peirce.

19. Proceedings of the 1st International Workshop on Peirce: A Conceptual Graphs Workbench.Ellis, G. and Levinson, R. editors. 1992 Jul 02; 1992; International Workshops on Peirce.

20. CEPTER (Concept space editor) [computer program]. Benneke A. Uni Hildesheim: 1995; HTTP. PC. Win NT, Win 95. http://www.med-informatik.uni-hildesheim.de/~benneke/Knowledge/knowledge.html.

21. NLM. The Unified Medical Language System. Bethesda. Natinal Library of Medicine. 1995; http://www.nlm.nih.gov/factsheets.dir/UMLS.html.

22. Eichmann D, McGregor T, and Danley D. Integrating Structured Databases Into the Web: The MORE System. COMNET. 27. 2. 1994; http://www.elsevier.nl/cgi-bin/query/WWW94/FinalProgramme.html?eichmann+mcgregor.

23. Conolly DW. HyperText Markup Language (HTML). Boston. World Wide Web Consortium. 1995; http://www.w3.org/hypertext/WWW/MarkUp/MarkUp.html.

24. Stein LD. The WWW Security FAQ. MIT. WWW. 1995; http://www-genome.wi.mit.edu/WWW/faqs/www-security-faq.

25. Varela C, Nekhayev P, Chanrasekharan P et al. Browsing Object-Oriented Databases over the Web. O'Reilly & Associates, Inc. 1995 Dec 11; 1995; p. 209

26. GSQL - a Mosaic-SQL gateway [computer program]. Ng J. 1993; http://www.ncsa.uiuc.edu/SDG/People/jason/pub/gsql/starthere.html.

27. The Lycos Catalog of the Internet 2.0.1.[computer program]. Pittsburgh, PA: Lycos, Inc. / Carnegie Mellon University; 1995; http://www.lycos.com.

28. Web Crawler [computer program]. Pinkerton B. America Online; 1995; http://webcrawler.com.

29. Bowman CM, Danzig PB, Hardy DR, Manber U, and Schwartz MF. The Harvest Information Discovery and Access System. 1994 Oct; Chicago, Illinois. 1994; p. 763; Proceedings of the Second International World Wide Web Conference.

30. Bowman CM, Danzig PB, Hardy DR, Manber U, and Schwartz MF. Harvest. 1996; http://harvest.cs.colorado.edu.

31. Bittorf A. DOIA FAQ. 1996; http://www.uni-erlangen.de/docs/derma/bilddb/doia_faq.htm.

32. Bittorf A, Krejci-Papa NC, and Diepgen TL. Dermatologic Online Image Atlas (DOIA). 1996;

http://www.uni-erlangen.de/docs/derma/bilddb/db.htm.

33. Bittorf A, Krejci-Papa NC, Ray T et al. Applications of the WWW: A Multicentered Dermatologic Image Database.Holzapfel, R. editor. Frauhofer Intitute for Computer Graphics; Darmstadt. 1995; p. 48;

34. Bittorf A, Krejci-Papa NC, and Diepgen TL. Development of a dermatological image atlas with worldwide access for the continuing education of physicians. Journal of Telemedicine and Telecare. 1. 1. pp.45-53. 1995; http://www.uni-erlangen.de/docs/derma/personen/asbittor/jtt/jtt.htm.

35. Bittorf A, Krejci-Papa NC, Hiller D, Bergmann M, and Diepgen TL. Development of an Image Data Bank which Aides in Clinical and Histopathological Differential Diagnosis as well as in Basic Research. Archives of Dermatologic Research. 287. 3-4. pp.378 1995;

36. Bittorf A, Krejci-Papa NC, and Diepgen TL. Storage and Retrieval of Digital Images in Dermatology. Skin Research and Technology. 1. pp.192-199. 1995;

37. Krejci-Papa NC, Bittorf A, and Diepgen TL. Online Reference Images for Dermatologists and Dermatopathologists. Dermatology Online Journal. 1. 1. pp.5 1995; http://matrix.ucdavis.edu/DOJvol1num1/atlas.html.

38. Bittorf A, Krejci-Papa NC, and Diepgen TL. Online Dienste: Referenz Bilddatenbank für Dermatologen. Eysenbach, G. editor. Springer; Heidelberg. 1996; Praxis und Computer. Noch nicht gedruckt.

39. Bittorf A, Krejci-Papa NC, and Diepgen TL. Teaching Resources for Dermatology on the WWW - Quiz System and Dynamic Lecture Scripts using a HTTP-Database Demon. J Am Med Informatics Assoc. 3. 1996;

40. Bittorf A, Fartasch M, and Diepgen TL. Auflösung und Kompression digitaler Bilder - Anforderungen an die Bildqualität in der Dermatologie. 1996 Sep 17; Bonn. 1996; GMDS. Vortrag.

41. Schosser RH, Sneiderman CA, and Pearson TG. How Dermatologists perceive CRT displays and siver halide prints of transperency-based images: A comparison study. Journal of Biological Photography. 62. 4. pp.135-137. 1994;

42. Sneiderman CA, Cookson JP, and Hood AF. Usage of Computer Graphic Images in Teaching Dermatology. Computerized Medical Imaging and Graphics. 16. 3. pp.151-152. 1992;

43. Perednia DA, Gaines JA, and Butruille TW. Comparison of the Clinical Informativeness of Photographs and Digital Imaging Media With Multiple-Choice Receiver Operating Characteristic Analysis. Archives of Dermatology. 131. pp.292-297. 1995;

44. Sneiderman CA, Schosser RH, and Pearson TG. A Comparison of JPEG and FIF Compression of Color Medical Images for Dermatology. Computerized Medical Imaging and Graphics. 18. 5. pp.339-342. 1994;

45. Poynton CA. High Definition Television and Desktop Computing. IEEE CES Tokyo Chaper; 1993 Jul 20; Tokyo. 1993; International Technical Workshop on Multimedia Tchnologies in the HDTV Age.

46. Poynton CA. Color Management Technology for Workstations. 1992 Sep 01; 1992; Sun Expo '92, Machester, UK. nonlinear RGB Space for HDTV: CCIR Recommendation 709; Color Management; Sun Color Management Framework.

47. Born G. Referenzhandbuch Dateiformate. Bonn. Addison Wesley. 1992;

48. Born G. Dateiformate Programmierhandbuch. Bonn. Addison Wesley. 1993;

49. Foley JD, Van Dam A, Feiner SK, Hughes JF, and Phillips RL. Einführung in die Computergraphik. Bonn, Paris, Reading u.a. Addison-Wesley. 1995;

50. Poynton CA. Frequently Asked Questions about Gamma. 1995; http://www.inforamp.com/~poynton/notes/colour_and_gamma/GammaFAQ.html.

51. Poynton CA. Poynton's Color FAQ. 1995; http://www.inforamp.com/~poynton/notes/colour_and_gamma/ColorFAQ.html.

52. Lilley C. Not Just Decoration: Quality Graphics for the Web. O'Reilly & Associates, Inc. 1995 Dec 11; 1995; p. 291

53. Bittorf A and Diepgen TL. Color correction for Digital Dermatologic Images. 1996 Oct; Washington. 1996; Proceedings des AMIA Annual Fall Symposium (formerly SCAMC). Abstract; accepted.

54. Meyer EA and Murray P. Phoenix Image Server. Elsevier; 1996 May 06; 1996; Fifth International World Wide Web Conference.